



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 165971

TO: Ralph J Gitomer
Location: 3d65 / 3c18
Art Unit: 1655
Friday, September 16, 2005

Case Serial Number: 10/650482

From: Noble Jarrell
Location: Biotech-Chem Library
Rem 1B71
Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

FOR OFFICIAL USE ONLY

Scientific and Technical Information Center

SEARCH REQUEST FORM

Requester's Full Name: Ralph C. Simon Examiner #: 69630 Date: 9/16/05
Art Unit: 1655 Phone Number: 2-0916 Serial Number: 10/650482
Location (Bldg/Room#): 3d65 (Mailbox #): 3013 Results Format Preferred (circle): PAPER DISK

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims; and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known.

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Seq 1-4 of Application
(ALL 6ADD34L)

STAFF USE ONLY

Searcher: Noble

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed: 9/16/05

Searcher Prep & Review Time: 16

Online Time: 10

Type of Search

2 NA Sequence (#)

2 AA Sequence (#)

Structure (#)

Bibliographic

Litigation

Fulltext

Other

Vendors and cost where applicable

STN Dialog

Questel/Orbit Lexis/Nexis

Westlaw WWW/Internet

In-house sequence systems

☒ Commercial ☐ Oligomer ☐ Score/Length
☐ Interference ☐ SPDI ☐ Encode/Transl
Other (specify)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:34:33 ; Search time 12251.5 Seconds
(without alignments)
11635.782 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2342

Sequence: 1 attttggcttcgtccac.....aaattatgtgctatctg 2942

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl:

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_on:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2934	99.7	2942	6	BD160691 Primer fo
2	2934	99.7	2942	6	AX884059 Sequence
3	2934	99.7	2942	9	AX027650 Homo sapi
4	2912.8	99.0	5264	9	BC065280 Homo sapi
5	2329	79.2	167079	9	AL606489 Human DNA
6	2329	79.2	328863	2	AL627202 Homo sapi
7	1365.6	46.4	5367	10	BC058078 Mus muscu
8	1303.4	44.3	166233	2	AC048369 Homo sapi
9	1175.4	40.0	215781	10	AC107842 Mus muscu
10	1137.4	38.7	220490	2	AC128396 Rattus no
11	1137.4	38.7	227054	2	AC098956 Rattus no
12	1137.4	38.7	248458	2	AC105827 Rattus no
13	1120.2	38.1	212261	2	AC119774 Rattus no
14	991.4	33.7	240836	2	AC094236 Rattus no
15	738	25.1	772	6	BD148615 Primer fo
16	738	25.1	772	6	AX868553 Sequence
17	618	21.0	166233	2	AC048369 Homo sapi
18	515	17.5	543	6	BD153658 Primer fo
19	515	17.5	543	6	AX873596 Sequence

c	20	413	14.0	1027	9	BC009873
	21	287.6	9.8	348	11	G53188
	22	265.8	9.0	561	6	AX341793 Sequence
	23	219	7.4	226	6	AX417281 Sequence
	24	219	7.4	226	6	AX977975 Sequence
	25	219	7.4	226	6	BD112834
	26	218	7.4	218	6	AX918015
	27	218	7.4	218	6	BD053548
	28	207.2	7.0	3704	10	BC006897
c	29	199	6.8	706	6	AX395830
	30	195	6.6	195	6	AX895150 Sequence
	31	195	6.6	195	6	BD030683 Sequence
	32	187.4	6.4	628	6	AX321354 Sequence
	33	61.8	2.1	2000	6	AX655393
c	34	61.2	2.1	125020	9	AF429315
	35	59.4	2.0	110000	14	AV318871_2
	36	57.4	2.0	125020	9	AF429315
c	37	51	1.7	7218	6	I66494
	38	50.4	1.7	110000	3	AC116305_3
	39	50	1.7	60360	5	CR391988
	40	48.6	1.7	110000	3	AC116957_2
c	41	48.6	1.7	162778	9	AC009559
	42	48	1.6	140982	5	BX842684
	43	47.6	1.6	245439	2	AC016186
	44	47.2	1.6	153445	2	AC143017
	45	46.4	1.6	8346	6	AX323714
	46	46.2	1.6	4001	6	CQ787366
c	47	46.2	1.6	167320	5	AL928691
c	48	46.2	1.6	177467	9	AC097510
	49	46	1.6	172637	10	AC123849
c	50	46	1.6	199698	2	BX004991
	51	45.8	1.6	149464	2	AC067800
	52	45.8	1.6	158811	9	AC004862
c	53	45.4	1.5	1141	6	AX083744
	54	45.4	1.5	228767	5	BX005261
	55	45	1.5	250029	3	AE014820
	56	44.8	1.5	4001	6	CQ787370
c	57	44.8	1.5	200953	2	AC130578
	58	44.8	1.5	253275	2	AC130147
	59	44.8	1.5	262008	2	AC129449
c	60	44.6	1.5	143096	2	CR387991
	61	44.2	1.5	39575	3	AC116925
c	62	44.2	1.5	108175	9	AC068206
c	63	44.2	1.5	153264	2	AC135216
c	64	44.2	1.5	162974	2	AC046166
c	65	44.2	1.5	167478	2	AC067723
	66	44.2	1.5	169571	9	AC020687
c	67	44.2	1.5	190102	9	AF003085
	68	44.2	1.5	191654	2	AC023176
	69	44.2	1.5	209152	2	AC015558
c	70	44	1.5	166850	2	AC012222
c	71	44	1.5	175346	2	AC130426
c	72	44	1.5	300684	1	AE017227
c	73	43.8	1.5	1051	9	CHRGIT
c	74	43.8	1.5	163495	5	BX321919
	75	43.8	1.5	178050	9	AC142334
	76	43.6	1.5	258144	2	BX927403
c	77	43.6	1.5	196685	2	AC092517
c	78	43.4	1.5	5986	6	AX323783
c	79	43.4	1.5	200418	9	AC074138
c	80	43.4	1.5	200688	2	AC073266
	81	43.2	1.5	6794	6	AX251873
	82	43.2	1.5	6794	6	AX344261
	83	43.2	1.5	6794	6	AX348652
c	84	43.2	1.5	11026	1	AE014120
c	85	43.2	1.5	155902	9	AP002852
	86	43.2	1.5	176734	2	CR377214
c	87	43.2	1.5	184332	2	AC011777
	88	43.2	1.5	215351	10	AL844212
c	89	43	1.5	37447	3	CEP11A5
c	90	43	1.5	122163	8	AC125793
c	91	42.8	1.5	137156	2	AC138846
c	92	42.8	1.5	148255	10	AC121524

BC009873	Homo sapi
G53188	SHGC-84724
AX341793	Sequence
AX417281	Sequence
AX977975	Sequence
BD112834	EST and e
AX918015	Sequence
BD053548	Sequence
BC006897	Mus muscu
AX395830	Sequence
AX895150	Sequence
BD030683	Sequence
AX321354	Sequence
AX655393	Sequence
AF429315	Homo sapi
Continuation (3 of	
AF429315	Homo sapi
I66494	Sequence 14
Continuation (4 of	
CR391988	Zebrafish
Continuation (3 of	
AC009559	Homo sapi
AX842684	Zebrafish
AC016186	Homo sapi
AC143017	Macaca mu
AX323714	Sequence
CQ787366	Sequence
AL928691	Zebrafish
AC097510	Homo sapi
AC123849	Mus muscu
BX004991	Danio rer
AC067800	Homo sapi
AC004862	Homo sapi
AX083744	Sequence
BX005261	Zebrafish
AE014820	Plasmodiu
CQ787370	Sequence
AC130578	Rattus no
AC130147	Rattus no
AC129449	Rattus no
CR387991	Danio rer
AC116925	Dicyste
AC068206	Homo sapi
AC135216	Bos tauru
AC046166	Homo sapi
AC067723	Homo sapi
AC020687	Homo sapi
AF003085	Homo sapi
AC023176	Homo sapi
AC015558	Homo sapi
AC012222	Homo sapi
AC130426	Homo sapi
AE017227	Mycobacte
M30947	Chimpanzee
BX321919	Zebrafish
AC142334	Pan trogl
BX927403	Danio rer
AC092517	Papio anu
AX323783	Sequence
AC074138	Homo sapi
AC073266	Homo sapi
AX251873	Sequence
AX344261	Sequence
AX348652	Sequence
AE014120	Buchnera
AP002852	Homo sapi
CR377214	Danio rer
AC011777	Homo sapi
AL844212	Mouse DNA
292830	Caenorhabdi
AC125793	Oryza sat
AC138846	Homo sapi
AC121524	Mus muscu


```
Db 1321 ACAGGCTAGCAAGGGCAAGATTTTACCCACCCCTGACCAAGATTAATGGCTACCAAGCCT 1380
Qy 1381 GGAGGAGAACACAGCCTTCTCCGATGGATCCAAAACACTGCAGAGATAACCCAAACACA 1440
Db 1381 GGAGGAGAACACAGCCTTCTCCGATGGATCCAAAACACTGCAGAGATAACCCAAACACA 1440
Qy 1441 GTTTGTTCTGCTCTCGAGACATTTCTGGAAACACCCAGGAATCCAAGAGAAAAT 1500
Db 1441 GTTTGTTCTGCTCTCGAGACATTTCTGGAAACACCCAGGAATCCAAGAGAAAAT 1500
Qy 1501 AGAATTATTAACACAGAGTTCCACTTGTCTTGGAGAGAGAGAGCCCTTCTGAGGCTG 1560
Db 1501 AGAATTATTAACACAGAGTTCCACTTGTCTTGGAGAGAGAGAGCCCTTCTGAGGCTG 1560
Qy 1561 TCCATCTAGTGAGATACCTATGGAAGAGGAGGCTGGAGAGGCGGAAATAGTGTAGTTGA 1620
Db 1561 TCCATCTAGTGAGATACCTATGGAAGAGGAGGCTGGAGAGGCGGAAATAGTGTAGTTGA 1620
Qy 1621 TTACTCATACCTAGAGGTGACCTTCCCAATTTCTGCGAGACACAGCTTGTGTAAACAACT 1680
Db 1621 TTACTCATACCTAGAGGTGACCTTCCCAATTTCTGCGAGACACAGCTTGTGTAAACAACT 1680
Qy 1681 GATAGATTATATTTTGGAGGTGCATCCAGTGACCTCGAAACAAAGTTCTGATCCAGAGG 1740
Db 1681 GATAGATTATATTTTGGAGGTGCATCCAGTGACCTCGAAACAAAGTTCTGATCCAGAGG 1740
Qy 1741 TGAGGATTGGGATGAGGAAGCTGAGGATGATGTTTGTAGTAGTATAGCTCACTGTGAGA 1800
Db 1741 TGAGGATTGGGATGAGGAGCTGAGGATGATGTTTGTAGTAGTATAGCTCACTGTGAGA 1800
Qy 1801 CTGAGACTTTGAAACAAGACCCCTGAAGGCTTCACTTTTGGAACTCTTCTGCGAGTGA 1860
Db 1801 CTGAGACTTTGAAACAAGACCCCTGAAGGCTTCACTTTTGGAACTCTTCTGCGAGTGA 1860
Qy 1861 TCCTTATAATCCCGAGAACTTTACAGCAACAAATTTGAGTAGTATAGCTCACTGTGAGA 1920
Db 1861 TCCTTATAATCCCGAGAACTTTACAGCAACAAATTTGAGTAGTATAGCTCACTGTGAGA 1920
Qy 1921 AGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA 1980
Db 1921 AGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA 1980
Qy 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGCTGGGAATC 2040
Db 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGCTGGGAATC 2040
Qy 2041 TAGTGCAAGTCAACGAGAGTCTCAACTGTGGAATCTCATTTCTGTAATCTGATGACCC 2100
Db 2041 TAGTGCAAGTCAACGAGAGTCTCAAACTGTGGAATCTCATTTCTGTAATCTGATGACCC 2100
Qy 2101 CTACACCCCTTTAAATTTTAAGGCTCCTTTTCAAACATCAGGGGAAAATGAGAAAGCTG 2160
Db 2101 CTACACCCCTTTAAATTTTAAGGCTCCTTTTCAAACATCAGGGGAAAATGAGAAAGCTG 2160
Qy 2161 TCGTGACTCAAAGACCCCATCTGAGTCAATTTGTGGCCATTTCTGAGTGCACACTTACT 2220
Db 2161 TCGTGACTCAAAGACCCCATCTGAGTCAATTTGTGGCCATTTCTGAGTGCACACTTACT 2220
Qy 2221 TTCTTGTAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCAGACTCGGTACAGCG 2280
Db 2221 TTCTTGTAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCAGACTCGGTACAGCG 2280
Qy 2281 TGAGCTTCTTCTGAGGAGACACACACATGTCAAAAGAAAAGGTAACCTTCTCTTGA 2340
Db 2281 TGAGCTTCTTCTGAGGAGACACACACATGTCAAAAGAAAAGGTAACCTTCTCTTGA 2340
Qy 2341 AGAAGTTACTCAGTATTATATAAGTGTGATGAGATGCGAAAGCAACATGGGAAGATT 2400
Db 2341 AGAAGTTACTCAGTATTATATAAGTGTGATGAGATGCGAAAGCAACATGGGAAGATT 2400
Qy 2401 TGCAAGGATGATGAGGTTCCAGAACGAAATTCAGAAACAGAAATGCTATTGGATA 2460
Db 2401 TGCAAGGATGATGAGGTTCCAGAACGAAATTCAGAAACAGAAATGCTATTGGATA 2460
```

```
Qy 2461 TTGCTTGACATTTTGAAACACAGAGAAAGAAATGTTTAAATAGACTCCAGGGAAACATGCTTCAA 2520
Db 2461 TTGCTTGACATTTTGAAACACAGAGAAAGAAATGTTTAAATAGACTCCAGGGAAACATGCTTCAA 2520
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Db 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Qy 2581 TACCTCTTACCTGAGAGGTGTCTTTTAAAAAACAATCTTGGCAGCTGTCTCTTTGACATTT 2640
Db 2581 TACCTCTTACCTGAGAGGTGTCTTTTAAAAAACAATCTTGGCAGCTGTCTCTTTGACATTT 2640
Qy 2641 TTTTCTTTAGAGAAATGTAACCTTGGAATCTAGTTAAATTTTCTTTTGGCAACATATCCC 2700
Db 2641 TTTTCTTTAGAGAAATGTAACCTTGGAATCTAGTTAAATTTTCTTTTGGCAACATATCCC 2700
Qy 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT 2760
Db 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT 2760
Qy 2761 CTAAATCTCTCTTTTGTGATTTAGTTGGATGTCTTTTAAATCTCTTTTGCCTGCATGAG 2820
Db 2761 CTAAATCTCTCTTTTGTGATTTAGTTGGATGTCTTTTAAATCTCTTTTGCCTGCATGAG 2820
Qy 2821 GTGGAAGGGGACCTTTTGTGATGTTGCACTTTTGCATCTTTCAAACTTATTTTCTTGAAA 2880
Db 2821 GTGGAAGGGGACCTTTTGTGATGTTGCACTTTTGCATCTTTCAAACTTATTTTCTTGAAA 2880
Qy 2881 ACAATATTATAGGCTTAAAGCCCATTTTCTTAATCTAAATATATGTTGTCCTATC 2940
Db 2881 ACAATATTATAGGCTTAAAGCCCATTTTCTTAATCTAAATATATGTTGTCCTATC 2940
Qy 2941 TG 2942
Db 2941 TG 2942
```

RESULT 2
AX884059

```
LOCUS 2942 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 18964 from Patent EP1074617.
ACCESSION AX884059
VERSION AX884059.1 GI:40038960
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayaashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesizing full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18964 07-FEB-2003;
RESEARCH Association for Biotechnology
FEATURES
Location/Qualifiers
1..2942
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
407..2548
/notes="unnamed protein product"
/codon_start=1
/db_xref="GI:40038961"
/translation="MEPTGGSRKRLPRAGFRFPWPPFPRRSQAGSSKFTPLPGPEN
SGNPTLSSAQPTVSRYTKLLSLQLLAPLGLLQKVLINSLQPLGCMFTRWLDKLEL
YSALRALKREKPAAPATQKLSLQSDSDSPVSPFLDWLEEGIHQWYPPDLKLEL
KAGSALDPAALQGLWGLVPSLQSLRYNRELSGSPSPFLNIQRIDDFSV
VYLINPSYLDQCFPRLEVSQNSGVQFOTLTPESSCLREDHCHPQLSABELIP
ASWQCPPLSTEGLEPIHLRMKRLFLQAKSGDLPDTPDODNGVHSLSEHSILRM
DPKHCNDNTPQFVPAAGDIPGNTQESTEKLELTTEVPLALEESPSGCSSEIPLM
EKPEPGEGRISVDYSYLEGLDLPISARPACSNKLDIYLGGASSDLETSDEPEDWDE
```

FEATURES
source

CDS

EAEDDGFDSLSLSDLEQDPPEGLHLWNSFCSVDPYNPQNFATQIOTAARIVPEPS
DSEKLSGSDLENSQSGLPETPEHSSBEDDWESSADBAESLKLWNSFCNSDDPY
NPLNPKAPQTSSENEKCRDSKTPSESIVASECHTLLSKVOLLGSESECRDPSVQ
RVLGSGRTHVKKVFLFLEVTEYISGDEDRKGPWEEFARDCRFQKRIQETEDA
IGYCLTFEHRERMFNRLOGTCKPGLNVLKQC*

ORIGIN		Query Match		Score 2934; DB 6; Length 2942;		Best Local Similarity 99.7%; Pred. No. 0;		Matches 2937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
QY	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGCCCTACCCAGTCCCTTCGCGGTATCGCGTTG	60						
DB	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGCCCTACCCAGTCCCTTCGCGGTATCGCGTTG	60						
QY	61	CTCAGGGGCTTTCAACCCCTCTGTAGTCGGAAACCATCGCGAGGCGGTGGGGGACT	120						
DB	61	CTCAGGGGCTTTTCAACCCCTCTGTAGTCGGAAACCATCGCGAGGCGGTGGGGGACT	120						
QY	121	CCTATCCATGTGTTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCGACGATGGA	180						
DB	121	CCTATCCATGTGTTGAAGCGTCGAGCGGACTAGGGAACCTCTTCCCGCCGACGATGGA	180						
QY	181	AGTCGCATCAGTCGCGGCTATTGCGCGGCTGTCTTCCCTGTGTCTGCGCGCGCTG	240						
DB	181	AGTCGCATCAGTCGCGGCTATTGCGCGGCTGTCTTCCCTGTGTCTGCGCGCGCTG	240						
QY	241	CGCGATTGCGTCCCTCTGTGTGCTTTCTGCTGGCTCGAGATCGGCTGGAGCGGAC	300						
DB	241	CGCGATTGCGTCCCTCTGTGTGCTTTTCTGCTGGCTCGAGATCGGCTGGAGCGGAC	300						
QY	301	GCACCGCTGGCAAGCGGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCCAGC	360						
DB	301	GCACCGCTGGCAAGCGGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCCAGC	360						
QY	361	CGCTGAGCGCCGCGCCCTACTGAGAGACTCTCAAGAAAAAGGAGATGGAGCCGGGAC	420						
DB	361	CGCTGAGCGCCGCGCCCTACTCTGAGAGACTCTCAAGAAAAAGGAGATGGAGCCGGGAC	420						
QY	421	AGGCGGATCGCGGAAACGCGCTTGGCCCTCGGCGGGCTTCCGGTCTTGGCCACCCCTTTT	480						
DB	421	AGGCGGATCGCGGAAACGCGCTTGGCCCTCGGCGGGCTTCCGGTCTTGGCCACCCCTTTT	480						
QY	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCCCGAGCGCTCTTGGCCCGGAAAACTC	540						
DB	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCCCGAGCGCTCTTGGCCCGGAAAACTC	540						
QY	541	CGGGAAACCCCACTGCTTTCTCTGCCAGCCCGAGACTCGGGTCAGTTACTGGACGAA	600						
DB	541	CGGGAAACCCCACTGCTTTCTCTGCCAGCCCGAGACTCGGGTCAGTTACTGGACGAA	600						
QY	601	ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCAGAGGTGCTAATTTGGAG	660						
DB	601	ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTTGCTTCAGAGGTGCTAATTTGGAG	660						
QY	661	CCAACTTTTCGCTGGAAATGTTTCCGACACAGATGGCTAGATTTTGTGAGTCTACAGCGC	720						
DB	661	CCAACTTTTCGCTGGAAATGTTTCCGACACAGATGGCTAGATTTTGTGAGTCTACAGCGC	720						
QY	721	CTGAGAGCCCTGAAAGGACGGGAGAAACACAGCGCCGCCACAGCGCAGAAATCTTTGAG	780						
DB	721	CTGAGAGCCCTGAAAGGACGGGAGAAACACAGCGCCGCCACAGCGCAGAAATCTTTGAG	780						
QY	781	TTTCGCTGACGCTCGACTCTCAGACCCCTCGGTCCAGTCCCTTTGATTGGCTAGAGGA	840						
DB	781	TTTCGCTGACGCTCGACTCTCAGACCCCTCGGTCCAGTCCCTTTGATTGGCTAGAGGA	840						
QY	841	GGGGATCCACTGGCATACTCGCCCGCCAGACCTAAAAATTGGAGCTTAAAGGCCAAGGGAAG	900						
DB	841	AGGGATCCACTGGCATACTCGCCCGCCAGACCTAAAAATTGGAGCTTAAAGGCCAAGGGAAG	900						
QY	901	TGCTTTGGACCCCTGACGACAGGCTTTCTCTTAGAGCAGCAGCTGTGGGGAGTGGAGCT	960						
DB	901	TGCTTTGGACCCCTGACGACAGGCTTTCTCTTAGAGCAGCAGCTGTGGGGAGTGGAGCT	960						

DB	901	TGCTTTGGACCCCTGACGACAGGCTTTCTCTTAGAGCAGCAGCTGTGGGGAGTGGAGCT	960
QY	961	GTTCCCGCAGTAGCCTTCAATCCCGCTGTACTCTAAACCGGAACTTGCCTCTTCGCCCTC	1020
DB	961	GTTCCCGCAGTAGCCTTCAATCCCGCTGTACTCTAAACCGGAACTTGCCTCTTCGCCCTC	1020
QY	1021	TGGGCTCTTAAACATTCAACGCGCATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAACCC	1080
DB	1021	TGGGCTCTTAAACATTCAACGCGCATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAACCC	1080
QY	1081	TTCTACTGTAAGTGTCTTTCAGGTAGAGTCAAGTATCAGAACTAGTGTGAAATAG	1140
DB	1081	TTCTACTGTAAGTGTCTTTCAGGTAGAGTCAAGTATCAGAACTAGTGTGAAATAG	1140
QY	1141	CGAGGTAGTGGGCTTCCAGACACTAACCCAGAGCAGCTGCGCTGAGAGAGGACCATTTG	1200
DB	1141	CGAGGTAGTGGGCTTCCAGACACTAACCCAGAGCAGCTGCGCTGAGAGAGGACCATTTG	1200
QY	1201	TCATCCCGCAGCGCTGAGTGCAGAACTCATTTCCGGCTCTGTTGGCAGGATGTCCACCTCT	1260
DB	1201	TCATCCCGCAGCGCTGAGTGCAGAACTCATTTCCGGCTCTGTTGGCAGGATGTCCACCTCT	1260
QY	1261	TTCTACGGAAGGCTTACAGAAATTCACATCTTTCGCAATGAAACGGCTGGAAATTCCTTCA	1320
DB	1261	TTCTACGGAAGGCTTACAGAAATTCACATCTTTCGCAATGAAACGGCTGGAAATTCCTTCA	1320
QY	1321	ACAGGCTAAACAAGGGGCAAGATTTTACCCACCCCTGACCAGGATAATGGCTACACAGCT	1380
DB	1321	ACAGGCTAGCAGGGGCAAGATTTTACCCACCCCTGACCAGGATAATGGCTACACAGCT	1380
QY	1381	GGAGGAGGAAACACAGCCCTTCTCGGATGGATCCAAAACACTGACAGAGATAACCAACACA	1440
DB	1381	GGAGGAGGAAACACAGCCCTTCTCGGATGGATCCAAAACACTGACAGAGATAACCAACACA	1440
QY	1441	GTTTGTCTGCTGCTGAGACATCTCTGGAACACCCAGGATCCACTGGAAGAAAAAT	1500
DB	1441	GTTTGTCTGCTGCTGAGACATCTCTGGAACACCCAGGATCCACTGGAAGAAAAAT	1500
QY	1501	AGAAATTAATACTACAGAGGTTCCACTTGTCTTGGAAAGAGAGAGCCCTCTGAGGGCTG	1560
DB	1501	AGAAATTAATACTACAGAGGTTCCACTTGTCTTGGAAAGAGAGAGCCCTCTGAGGGCTG	1560
QY	1561	TCCATCTAGTGAGATCCTATGGAAGAGGAGCCCTGGAAGGGCCGAATAAGTGTAGTTGA	1620
DB	1561	TCCATCTAGTGAGATCCTATGGAAGAGGAGCCCTGGAAGGGCCGAATAAGTGTAGTTGA	1620
QY	1621	TTACTCATACCTAGAGGTGACCTTCCCAATTTCTGCCAGACCCAGCTTGTAGTAACT	1680
DB	1621	TTACTCATACCTAGAGGTGACCTTCCCAATTTCTGCCAGACCCAGCTTGTAGTAACT	1680
QY	1681	GATAGATTATATTTTGGAGGTGCATCCAGTGACCTGGAAACAAAGTTCTGATCCAGAGG	1740
DB	1681	GATAGATTATATTTTGGAGGTGCATCCAGTGACCTGGAAACAAAGTTCTGATCCAGAGG	1740
QY	1741	TGAGGATTGGGATGAGGAAGCTGAGGATGATGTTTGTAGTGTAGTACGTCACTGTGAGA	1800
DB	1741	TGAGGATTGGGATGAGGAAGCTGAGGATGATGTTTGTAGTGTAGTACGTCACTGTGAGA	1800
QY	1801	CTCAGACTTGAACAAAGCCCTGAGGGCTTCACTTTGGAACTCTTCTGCACTGTAGA	1860
DB	1801	CTCAGACTTGAACAAAGCCCTGAGGGCTTCACTTTTGGAACTCTTCTGCACTGTAGA	1860
QY	1861	TCCTTATAATCCCGAGAACTTTTACAGCAAACTTTCAGACTGCTGCCAGAAATTTGCTCTGA	1920
DB	1861	TCCTTATAATCCCGAGAACTTTTACAGCAAACTTTCAGACTGCTGCCAGAAATTTGCTCTGA	1920
QY	1921	AGAGCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGTATCTAGAGAAATTCCTCCCA	1980
DB	1921	AGAGCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGTATCTAGAGAAATTCCTCCCA	1980
QY	1981	GTCTGGAGAGCTTCTCAGACCCCTGAGACTAGTTCTTGGGGAGGAGATGACTGGGAATC	2040
DB	1981	GTCTGGAGAGCTTCTCAGACCCCTGAGACTAGTTCTTGGGGAGGAGATGACTGGGAATC	2040

Qy	2041	TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATTTCTGTAAATCTGATGACCC	2100
Db	2041	TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATTTCTGTAAATCTGATGACCC	2100
Qy	2101	CTACAACCCCTTAAATTTTAAGGCTCTCTTTTCAAAACATCAGGGGAAAATGAGAAAGCTG	2160
Db	2101	CTACAACCCCTTAAATTTTAAGGCTCTCTTTTCAAAACATCAGGGGAAAATGAGAAAGCTG	2160
Qy	2161	TCGTGACTCAAGACCCCATCTGAGTGCATTTCTGTGGCCATTTCTCAGTGTACACCTTACT	2220
Db	2161	TCGTGACTCAAGACCCCATCTGAGTGCATTTCTGTGGCCATTTCTCAGTGTACACCTTACT	2220
Qy	2221	TTCTTTGAAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Db	2221	TTCTTTGAAGGTGAGCTGTTGGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Qy	2281	TGAGCTTCTTCTCGGAGAGACACACACATGTCAAAAAGAAAAGGTAACTCTTCCTGA	2340
Db	2281	TGAGCTTCTTCTCGGAGAGACACACACATGTCAAAAAGAAAAGGTAACTCTTCCTGA	2340
Qy	2341	AGAACTTACTGAGTATTATATAAGTGTGTAGAGTGCAGAAAGGACCATGGGAAGATT	2400
Db	2341	AGAACTTACTGAGTATTATATAAGTGTGTAGAGTGCAGAAAGGACCATGGGAAGATT	2400
Qy	2401	TGCAAGGGATGGATGCAGGTTCCAGAAACGAATTCAGAGAAACAGAGATGCTATTGGATA	2460
Db	2401	TGCAAGGGATGGATGCAGGTTCCAGAAACGAATTCAGAGAAACAGAGATGCTATTGGATA	2460
Qy	2461	TTGCTTGACATTTGAAACACAGAGAAAGATTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Db	2461	TTGCTTGACATTTGAAACACAGAGAAAGATTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Qy	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTGGCAGCTGTAGTCTGAGTGTAGCTATCAC	2580
Db	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTGGCAGCTGTAGTCTGAGTGTAGCTATCAC	2580
Qy	2581	TACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATTT	2640
Db	2581	TACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATTT	2640
Qy	2641	TTTTTTTTAGAGGAATGTAATCTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Db	2641	TTTTTTTTAGAGGAATGTAATCTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Qy	2701	ACTCAGAAACATTTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAAGTGTGATTT	2760
Db	2701	ACTCAGAAACATTTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAAGTGTGATTT	2760
Qy	2761	CTAATCTCTCCCTTTTGTATTTAGTGTGATGTGCTTTTAAATGTCTTTGCTGCATGAG	2820
Db	2761	CTAATCTCTCCCTTTTGTATTTAGTGTGATGTGCTTTTAAATGTCTTTGCTGCATGAG	2820
Qy	2821	GTGAAAGGGGACCTTTTGTGATTTGATTTGACATTTTCAAACTTATTTCTTGGA	2880
Db	2821	GTGAAAGGGGACCTTTTGTGATTTGATTTGACATTTTCAAACTTATTTCTTGGA	2880
Qy	2881	ACAATATTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTGTGCTTATC	2940
Db	2881	ACAATATTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATATGTGTGCTTATC	2940
Qy	2941	TG 2942	
Db	2941	TG 2942	

RESULT 3

AK027650

LOCUS

AK027650 Homo sapiens cDNA FLJ14744 fis, clone NT2RP3002770. linear PRI 30-JAN-2004

ACCESSION

AK027650

VERSION

AK027650.1

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,

Wakamatsu, A., Hayashi, K., Sato, H., Negai, K., Kimura, K., Makita, H.,

Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T.,

Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,

Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,

Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Kanda, K., Kanda, K., Yokoi, T., Furuya, T.,

Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,

Tanikawa, K., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,

Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,

Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

Yosida, M., Hotuta, T., Kusan, J., Kanehori, K., Takahashi, Fujii, A.,

Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,

Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,

Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,

Shiohata, N., Sano, S., Mori, S., Mori, S., Momiya, H., Satoh, N., Takami, S.,

Teraohima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,

Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,

Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,

Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,

Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,

Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,

Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,

Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,

Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.

Complete sequencing and characterization of 21,243 full-length

human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)

14702039

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

2

Isogai, T., Ota, T., Hayaashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,

Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K.,

Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,

Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.

NEDO human cDNA sequencing project

TITLE

JOURNAL

REFERENCE

AUTHORS

3

(bases 1 to 2942)

Isogai, T. and Otsuki, T.

TITLE

JOURNAL

REFERENCE

AUTHORS

1

Location/Qualifiers

1. 2942

/organism="Homo sapiens"

/mol_type="mrna"

/db_xref="taxon:9606"

/clone="NT2RP3002770"

/cell_line="NT2"

/cell_type="teratocarcinoma"

/notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal

precursor cells after 2-weeks retinoic acid (RA)

induction."

407. .2548

/notes="unnamed protein product"

CDS

/codon_start=1									
/protein_id="BAB55266.1"									
/db_xref="GI:14042485"									
/translation="MEPRTGSRKRLGPRAGFRWPPFPFRSRQAGSKFPTPLGPN									
SGNPTLLSQAQPETRVTKLLSLLQAPLGLQKVLINSQLFGGMFPTRWLDFAGV									
YSALRALGRKPAAPTAKSLSLQSDSPSVTSPLDMLBEGHWHQYSPPDLKJEL									
KAKGSALPAQAFLLEQQLAGVELLPSSLRSLYSNRELGSFSPGLNIQRIDFVS									
VSYLLNPVLDGFCPLRELYVQNSDGNVWGFQTLTPRESSCLREDHCHPOPLSABLIP									
ASWQCCPLSTEGLEPIHILRMKRLFELQAASKGQDLPTPDQDNGYHSEELSLRM									
DPKCRDNPQFVPAAGDI PGNTQBSTEEKIELLTTEVPLALBESPSGSPSEI PM									
EKBPGEIRISVVDYSLEGDLPISARPACSNKLIDYILGGAASDLETSSPEGEWDDE									
EAEKDFDSDLSLSDLEQPEGLHLWNSFSDVPYNPONTATIOTAAARIVPEPS									
DSEKDLGSKDLSNSQSGSLPETPEHSGSBEEDWESSADEAESIKLWNSFCNSDDPY									
NPLNPKAPQTSGENEKCRDSKTPBSI VAI SECHTLLSKVOLLGSOBSEC PDSVQ									
RDVLSGGRHTWVKRKKVFLBEVTVYISGEDRKG PWEERFARDGCRPQKRIQETEDA									
IGYCLTPEHREMENRLQGTCPKGLNLVKQC"									
ORIGIN									
Query Match 99.7%; Score 2934; DB 9; Length 2942;									
Best Local Similarity 99.8%; Pred. No. 0;									
Matches 2937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;									
Qy	1	ATTTTGGGCTTCGCTTCACACGACACAGCGGCTTACCCAGTCTTCCTCGGTATCGCGTTG	60						
Db	1	ATTTTGGGCTTCGCTTCACCCGACACAGCGGCTTACCCAGTCTTCCTCGGTATCGCGTTG	60						
Qy	61	CTCAGGGCTTTTCAACCTCTGTGTCAGTCGGAACACCATCGCCGAGGCGCTGGGGGACT	120						
Db	61	CTCAGGGGCTTTTCAACCTCTGTGTCAGTCGGAACACCATCGCCGAGGCGCTGGGGGGACT	120						
Qy	121	CTATTCATGTGTTGAAGCGTCGAGCGGACTTAGGGAACCTCTTCCCGCCGACGAGATGGA	180						
Db	121	CTATTCATGTGTTGAAGGTCGAGCGGACTTAGGGAACCTCTTCCCGCCGACGAGATGGA	180						
Qy	181	AGTCGATCAGTCGCGCGCTATGTCGCGGCGTGTCTTCCTCGTGTCTGCGCCCGCTG	240						
Db	181	AGTCGATCAGTCGCGCGCTATGTCGCGGCGTGTCTTCCTCGTGTCTGCGCCCGCTG	240						
Qy	241	CGCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAAGATCGGCTGGAGCAGCGAC	300						
Db	241	CGCATTCGCTGCCCTCTGTGGCTTTTCTGCTGGCTCGAAGATCGGCTGGAGCAGCGAC	300						
Qy	301	GCCACCGCTGGCAAGCGCGAGACTCTGTAGGCTTCTTCGGAATCCCGTGCACCTCCAGC	360						
Db	301	GCCACCGCTGGCAAGCGCGAGACTCTGTAGGCTTCTTCGGAATCCCGTGCACCTCCAGC	360						
Qy	361	CGCTAGCGCCGCGGCTTACTGTAGAGACTGTCAAGAAAAGGAGATGGAGCGGGGAC	420						
Db	361	CGCTAGCGCCGCGGCTTACTGTAGAGACTGTCAAGAAAAGGAGATGGAGCGGGGAC	420						
Qy	421	AGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGGGCTTCCGGTTCGGCCACCCCTTTT	480						
Db	421	AGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGGGCTTCCGGTTCGGCCACCCCTTTT	480						
Qy	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCCTCGA CGCTCTTGGCCCGGAAAACTC	540						
Db	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCCTCGA CGCTCTTGGCCCGGAAAACTC	540						
Qy	541	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTACTGACGAA	600						
Db	541	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTACTGACGAA	600						
Qy	601	ACTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTTGCTTCAGAAAGTGTCTAATTTGGAG	660						
Db	601	ACTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTTGCTTCAGAAAGTGTCTAATTTGGAG	660						
Qy	661	CCAACTTTTCGTTGGAAATGTTTCCGACAGATGGCTAGATTTTGTGCTGAGTCTACAGCGC	720						
Db	661	CCAACTTTTCGTTGGAAATGTTTCCGACAGATGGCTAGATTTTGTGCTGAGTCTACAGCGC	720						
Qy	721	CCTGAGAGCCCTGAAGGAGCGGAGAAACACGCGCCGCCACAGCGCAGAAATCTTTGAG	780						
Db	721	CCTGAGAGCCCTGAAGGAGCGGAGAAACACGCGCCGCCACAGCGCAGAAATCTTTGAG	780						

Qy	781	TTTCGCTCGAGCTCGACTCTCTCAGACCCCTCGGTCCAGCTCCCTTGATTGGCTAGAGGA	840
Db	781	TTTCGCTCGAGCTCGACTCTCTCAGACCCCTCGGTCCAGCTCCCTTGATTGGCTAGAGGA	840
Qy	841	GGGATCCATCGGCAATATCTCGCCCCCAGACCTTAAATTTGGAGCTTAAGGCAAGGAAG	900
Db	841	AGGATCCATCGGCAATATCTCGCCCCCAGACCTTAAATTTGGAGCTTAAGGCAAGGAAG	900
Qy	901	TGCTTTGGAACCTTGACACAGGCTTTCTTTAGACACAGCTGTGGGAGTGGAGCT	960
Db	901	TGCTTTGGAACCTTGACACAGGCTTTCTTTAGACACAGCTGTGGGAGTGGAGCT	960
Qy	961	GTTCGCCAGTAGCTTCAATCCCGTCTGTACTTAAACCGGGAATCTTGGCTCTTCGCCCTC	1020
Db	961	GTTCGCCAGTAGCTTCAATCCCGTCTGTACTTAAACCGGGAATCTTGGCTCTTCGCCCTC	1020
Qy	1021	TGGGCTCTTAAACATTAACGCAATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAAACC	1080
Db	1021	TGGGCTCTTAAACATTAACGCAATAGACAAATTTTCAGTGTGGTATCTTATTTGCTGAAACC	1080
Qy	1081	TTCCCTACCTCGACTGCTTCCAGACACTTAAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATG	1140
Db	1081	TTCCCTACCTCGACTGCTTCCAGACACTTAAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATG	1140
Qy	1141	CGAGGTAGTTCGGCTTCCAGACACTTAAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATG	1200
Db	1141	CGAGGTAGTTCGGCTTCCAGACACTTAAACCCAGAGAGCAGCTGCCTGAGAGAGGACCAATG	1200
Qy	1201	TCATTCGAGCGCTGAGTGAGAACTCATTTCCGGCTCTGTCGGCAGGGATGTCCACCTCT	1260
Db	1201	TCATTCGAGCGCTGAGTGAGAACTCATTTCCGGCTCTGTCGGCAGGGATGTCCACCTCT	1260
Qy	1261	TTCTACCGAAGGCGCTACACAGAAATTCACCATTTCCGATGAAACGGCTGGAATTCCTTCA	1320
Db	1261	TTCTACCGAAGGCGCTACACAGAAATTTCCGATGAAACGGCTGGAATTCCTTCA	1320
Qy	1321	ACAGGCTTAAACAAAGGGGCAAGATTTTACCCACCCCTTGAC CAGGATAATGGCTACCA CAGCCT	1380
Db	1321	ACAGGCTTAGCAAGGGGCAAGATTTTACCCACCCCTTGAC CAGGATAATGGCTACCA CAGCCT	1380
Qy	1381	GGAGGAGAAACACAGCCTTCTCGGATGGATCCAAACACTGCGAGATATACCCCAACACA	1440
Db	1381	GGAGGAGAAACACAGCCTTCTCGGATGGATCCAAACACTGCGAGATATACCCCAACACA	1440
Qy	1441	GTTCGCTGCTGCTGAGACATTTCTGGAACACCCAGGAAATCCACTGAAGAAAAAAT	1500
Db	1441	GTTCGCTGCTGCTGAGACATTTCTGGAACACCCAGGAAATCCACTGAAGAAAAAAT	1500
Qy	1501	AGAAATTAATTAACCTACAGAGGTTCCACTTGTGTTGGAAGAGAGAGCCCTTCTGAGGGCTG	1560
Db	1501	AGAAATTAATTAACCTACAGAGGTTCCACTTGTGTTGGAAGAGAGAGCCCTTCTGAGGGCTG	1560
Qy	1561	TCCATCTAGTGAGATACCTATGGAAGAGGAGCTGGAGAGGCGGATAGTGTAGTTGA	1620
Db	1561	TCCATCTAGTGAGATACCTATGGAAGAGGAGCTGGAGAGGCGGATAGTGTAGTTGA	1620
Qy	1621	TTACTCATACCTAGAAGTGACCTTCCCAATTTCTGCGAGACAGCTTGTAGTAACAAACT	1680
Db	1621	TTACTCATACCTAGAAGTGACCTTCCCAATTTCTGCGAGACAGCTTGTAGTAACAAACT	1680
Qy	1681	GATAGATTATATTTTGGGAGGTGCATCCAGTGACTCGGAAACAAAGTTCTGATCCAGAAAG	1740
Db	1681	GATAGATTATATTTTGGGAGGTGCATCCAGTGACTCGGAAACAAAGTTCTGATCCAGAAAG	1740
Qy	1741	TGAGGATTTGGATGAGAACTGAGGATGATGGTTTGTAGTGATAGTACCTGTGAG	1800
Db	1741	TGAGGATTTGGATGAGAACTGAGGATGATGGTTTGTAGTGATAGTACCTGTGAG	1800
Qy	1801	CTCAGACCTTGAAACAGACCCCTGAAGGGCTTCACTTTTGGAACTCTTTTCTGCACTGTAGA	1860
Db	1801	CTCAGACCTTGAAACAGACCCCTGAAGGGCTTCACTTTTGGAACTCTTTTCTGCACTGTAGA	1860

Qy	1861	TCCTTATAATCCCCAGAACCTTTACAGCAACAATTCAGACTGCTGCCAGAAATTTGTTCTCTGA	1920
Db	1861	TCCTTATAATCCCCAGAACCTTTACAGCAACAATTCAGACTGCTGCCAGAAATTTGTTCTCTGA	1920
Qy	1921	AGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA	1980
Db	1921	AGAGCCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCCA	1980
Qy	1981	GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTTGGGGAGGAAGATGACTGGGAATC	2040
Db	1981	GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTTGGGGAGGAAGATGACTGGGAATC	2040
Qy	2041	TAGTGAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCAATCTGTAAATTCATGATGACCC	2100
Db	2041	TAGTGAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCAATCTGTAAATTCATGATGACCC	2100
Qy	2101	CTACAAACCCCTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAATGAGAAAGGCTG	2160
Db	2101	CTACAAACCCCTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAATGAGAAAGGCTG	2160
Qy	2161	TCGTGACTCAAGACCCCTCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT	2220
Db	2161	TCGTGACTCAAGACCCCTCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT	2220
Qy	2221	TTCTTGAAGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Db	2221	TTCTTGAAGTGCAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG	2280
Qy	2281	TGAGTTCTTCTGGAGGAAGACACACATGTCAAAAGAAAAGTAACTTCTCTTGA	2340
Db	2281	TGAGTTCTTCTGGAGGAAGACACACATGTCAAAAGAAAAGTAACTTCTCTTGA	2340
Qy	2341	AGAAGTTACTGAGTATTATATAAGTGTGTAGAGATCGCAAGGACCATCGGGAAGAATT	2400
Db	2341	AGAAGTTACTGAGTATTATATAAGTGTGTAGAGATCGCAAGGACCATCGGGAAGAATT	2400
Qy	2401	TGCAAGGATGATGTCAGGTTCCAGAAACGAATTCAGAAACAGAAATGCTATTGATA	2460
Db	2401	TGCAAGGATGATGTCAGGTTCCAGAAACGAATTCAGAAACAGAAATGCTATTGATA	2460
Qy	2461	TTGCTTGACATTTGAACACAGAGAAAGATTTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Db	2461	TTGCTTGACATTTGAACACAGAGAAAGATTTTAAATAGACTCCAGGGAACATGCTTCAA	2520
Qy	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCTAGCTAGCATACAC	2580
Db	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCTAGCTAGCATACAC	2580
Qy	2581	TACCTCTTACCTGAGAGGTGCTTTTAAAAACAAATCTTGGCAGCTGTCTTTGACATTT	2640
Db	2581	TACCTCTTACCTGAGAGGTGCTTTTAAAAACAAATCTTGGCAGCTGTCTTTGACATTT	2640
Qy	2641	TTTTTTTTTAGAGGAATGTAATCTTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Db	2641	TTTTTTTTTAGAGGAATGTAATCTTGGATCTAGTTAAATTTTTTTTTTTTGGCAATATCCC	2700
Qy	2701	ACTCAGAAACATTCAGGTTTCAAGCCGCTGATTAATGAAGGATGAAGTGTGATTT	2760
Db	2701	ACTCAGAAACATTCAGGTTTCAAGCCGCTGATTAATGAAGGATGAAGTGTGATTT	2760
Qy	2761	CTAATCTCTCCCTTTTGTATTTAGTTGGATGTGCTTTTAAATGTCTTGTCTGCATGAG	2820
Db	2761	CTAATCTCTCCCTTTTGTATTTAGTTGGATGTGCTTTTAAATGTCTTGTCTGCATGAG	2820
Qy	2821	GTGGAAGGGACCTTTTGTGATTTGATTTGACATTTTCAAACTTATTTCTTGGAAA	2880
Db	2821	GTGGAAGGGACCTTTTGTGATTTGATTTGACATTTTCAAACTTATTTCTTGGAAA	2880
Qy	2881	ACAAATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATGTGTCCTATC	2940
Db	2881	ACAAATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATGTGTCCTATC	2940
Qy	2941	TG 2942	

Db	2941	TG 2942	
RESULT 4			
LOCUS	BC065280		
DEFINITION	Homo sapiens cDNA clone MGC:74824 IMAGE:6172811, complete cds.		
ACCESSION	BC065280		
VERSION	BC065280.1	GI:40807094	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 5264)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.W., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 5264)		
AUTHORS	Director MGC Project.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-JAN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabs-rc@mail.nih.gov Tissue Procurement: AFCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisn.nih.gov/ Contact: nisc.mgc@nih.gov Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, J., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
FEATURES	Location/Qualifiers		
source	1. 5264 /organism="Homo sapiens"		
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAC Plate: 140 Row: p Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 41282220.			

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:74824 IMAGE:6172811"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
381..2522
CDS
/codon_start=1
/product="Unknown (protein for MGC:74824)"
/protein_id="AAH65280.1"
/db_xref="GI:40807095"
/translation="MEPGTGGSRKRLGPRAGFRPMPFFRRSQSSGSKFTPLPGEN
SGNPTLLSSADPETRVSYWTKLLSOLLAPLQKLVILWSQLFGMPKPRWLDFAGV
YSALRALGRKREPAAPTAQKLSLSIQLDSDPSVTSPLDWEELHMYOYSPDLKLEL
KAKGALDPAQAQALFLEQLMGVELLPSLSQRLYSNRELSSPSGLINIORINFVS
VSYLLNPFLDCEPLEVSYQNSDNGENVGFQITLPESSCLREDHCHPQPSAELPL
ASWQCPLSTEGLEPIIHILMKRLEFLQQAQKQDLPTPDQNGYSLIEEHEELLRM
DPKHCRDNPOTFPAAGDI PGNTQSTPEKEI ELITTEVPLALEESESSEGPCSEI PM
EKEPEGRI SVVDYSYLRGDLPI SARPAKSNKLDYILGGASDLSTSSDPEDGWDPE
BAEDDGPDSLSLSDLEODPEGLHLMNSFCSDVPYNPQNTAT IOTAA RIVPEEPS
DSEKDLGSKSLDENSQSGLPETPEHSGREDDWESSADEASIKLWNSFCNSDDPY
NPLNKAFFQTSGENEKCRDSETPSESI VALSECHTLLSKVQLLGSQSECDPSVQ
RDVLSGGHTRVKKKXVFLBEVTEYISGDEDRKPWEEFARDGCRPQKRQIQTEDA
IGYCLTFEHRERMFNRLQGTCTFKGLNVLKQC"
```

ORIGIN

Query Match	99.0%;	Score 2912.8;	DB 9;	Length 5264;
Best Local Similarity	99.9%;	Pred. No. 0;	Mismatches 2;	Indels 0;
Matches 2914;	Conservative 0;	Gaps 0;		
QY	27	AGCCGGCCTACCCAGTCCCTTCGCGTATCGCGTGTCTCAGGGGCTTTTCAACCCCTCTGTCA	86	
DB	1	AGCCGGCCTACCCAGTCCCTTCGCGTATCGCGTGTCTCAGGGGCTTTTCAACCCCTCTGTCA	60	
QY	87	GTGCGAAAACCATCGCCGAGGCCGTGGGGGAGCTCCTATCCATGCTGTGAAGCGTCCGAG	146	
DB	61	GTGCGAAAACCATCGCCGAGGCCGTGGGGGAGCTCCTATCCATGCTGTGAAGCGTCCGAG	120	
QY	147	CCGACTAGGGAACCTCTTCCCGCCAGGATGGAAGTGCATCAGTCCGCCCTATTTCG	206	
DB	121	CCGACTAGGGAACCTCTTCCCGCCAGGATGGAAGTGCATCAGTCCGCCCTATTTCG	180	
QY	207	CGGGCTGTCTTCCCTGTGTTCTGCGCCCGCTGCGCATTCGCTGCCCTCTGTGGCTTT	266	
DB	181	CGGGCTGTCTTCCCTGTGTTCTGCGCCCGCTGCGCATTCGCTGCCCTCTGTGGCTTT	240	
QY	267	TCTGTGGCTCGAAGATCGGCTTGAGCAGCGAGCCGACCGCTGGGCAAGCGCGAGCTC	326	
DB	241	TCTGTGGCTCGAAGATCGGCTTGAGCAGCGAGCCGACCGCTGGGCAAGCGCGAGCTC	300	
QY	327	TGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCCGCGGCCCTACCTGAG	386	
DB	301	TGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCCGCGGCCCTACCTGAG	360	
QY	387	AGACTGTCAAGAAAAGAGATGGAGCCGGGACAGCGGATTCGCGAAAACGGCTTGGCC	446	
DB	361	AGACTGTCAAGAAAAGAGATGGAGCCGGGACAGCGGATTCGCGAAAACGGCTTGGCC	420	
QY	447	CTCGGGCGGGCTTCGGGTTCTGGCCACCCTTTTTCCTTCGGGATTCGCAAGAGGCTTT	506	
DB	421	CTCGGGCGGGCTTCGGGTTCTGGCCACCCTTTTTCCTTCGGGATTCGCAAGAGGCTTT	480	
QY	507	CTAAGTCCCGACGCTCTTGGCCCGGAAACTCCGGGAAACCCACACTGCTTTCCTCTG	566	
DB	481	CTAAGTCCCGACGCTCTTGGCCCGGAAACTCCGGGAAACCCACACTGCTTTCCTCTG	540	
QY	567	CCAGCCCGGAGACTCGGGTCAAGTACAGCAAACTGCTCTCCAGCTCTCTTCGCGCGC	626	
DB	541	CCAGCCCGGAGACTCGGGTCAAGTACAGCAAACTGCTCTCCAGCTCTCTTCGCGCGC	600	
QY	627	TCCCGGATTCGTACAGAGGTGCTTAATTGGAGCCAACTTTTCGGTGAATGTTCCGA	686	

DB	601	TCCCGGATTGCTTTACAGAGGTGCTAAATTTGGAGCCAACTTTTCGGTGAATGTTCCGA	660	
QY	687	CCAGATCGCTAGATTTTGTCTGGAGTCTACAGCGCCCTGAGAGCCCTCAAGGACGAGAGA	746	
DB	661	CCAGATCGCTAGATTTTGTCTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGAGCGGAGA	720	
QY	747	AACCAGCGCCGCCACAGCGCAGAAATCTTTTGAGTTCGCTGCAGCTCGACTCTCTCAGACC	806	
DB	721	AACCAGCGCCGCCACAGCGCAGAAATCTTTTGAGTTCGCTGCAGCTCGACTCTCTCAGACC	780	
QY	807	CCTCGGTACAGTCCCTCTGATTTGGCTAGAGGAGGATCCACTGGAATACTTCGCCCC	866	
DB	781	CCTCGGTACAGTCCCTCTGATTTGGCTAGAGGAGGATCCACTGGAATACTTCGCCCC	840	
QY	867	CAGACCTAAAATTTGGAGCTTAAAGCCAAAGCAAGTCTTTGGACCTTCGACGACAGGCTT	926	
DB	841	CAGACCTAAAATTTGGAGCTTAAAGCCAAAGCAAGTCTTTGGACCTTCGACGACAGGCTT	900	
QY	927	TTCTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCTGTTTGGCCAGTAGCTTTCAATCCCGTC	986	
DB	901	TTCTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCTGTTTGGCCAGTAGCTTTCAATCCCGTC	960	
QY	987	TGTACTCTAAACCGGAACTTTGGCTTCTCGCCCTCTGGCCCTCTAAACATTTCAACGCAATAG	1046	
DB	961	TGTACTCTAAACCGGAACTTTGGCTTCTCGCCCTCTGGCCCTCTAAACATTTCAACGCAATAG	1020	
QY	1047	ACAATTTCACTGTGTTATCTTATTTGCTGAACCTCTCTACCTGGAGCTGCTTCTTAGGC	1106	
DB	1021	ACAATTTCACTGTGTTATCTTATTTGCTGAACCTCTCTACCTGGAGCTGCTTCTTAGGC	1080	
QY	1107	TAGAAGTCAGCTATCAGAACAGTGATGGAATAATGCGAGGTAGTCGGCTTCAGACACTAA	1166	
DB	1081	TAGAAGTCAGCTATCAGAACAGTGATGGAATAATGCGAGGTAGTCGGCTTCAGACACTAA	1140	
QY	1167	CCCAGAGAGAGCTGCTGCTGAGAGAGCAACATTTGTCTATCCCAGCCGCTGAGTGCAAGAC	1226	
DB	1141	CCCAGAGAGAGCTGCTGCTGAGAGAGCAACATTTGTCTATCCCAGCCGCTGAGTGCAAGAC	1200	
QY	1227	TCATTTCCGGCTGTGGCAGGGATGTCACCTCTTTCTACGGAAGGCTTACCAAGAAATTC	1286	
DB	1201	TCATTTCCGGCTGTGGCAGGGATGTCACCTCTTTCTACGGAAGGCTTACCAAGAAATTC	1260	
QY	1287	ACCATTTCTGCATGAAAACGGCTGGAATTCCTTCAACAGGCTTAAACAGGGGCAAGATTAC	1346	
DB	1261	ACCATTTCTGCATGAAAACGGCTGGAATTCCTTCAACAGGCTTAAACAGGGGCAAGATTAC	1320	
QY	1347	CCACCCCTGACCAAGTAATGCGCTACCAAGCTGGAGGAGGACACAGGCTTCTCCGGA	1406	
DB	1321	CCACCCCTGACCAAGTAATGCGCTACCAAGCTGGAGGAGGACACAGGCTTCTCCGGA	1380	
QY	1407	TGGATCCAAAACACTGACAGAGATAAACCCAAACACAGTTTGTTCCTGCTGCTGGAGACATTC	1466	
DB	1381	TGGATCCAAAACACTGACAGAGATAAACCCAAACACAGTTTGTTCCTGCTGCTGGAGACATTC	1440	
QY	1467	CTGAAAACACCCAGGAATCCACTGGAAGAAAAATAGAAATTTATTAACCTACAGAGGTTCCAC	1526	
DB	1441	CTGAAAACACCCAGGAATCCACTGGAAGAAAAATAGAAATTTATTAACCTACAGAGGTTCCAC	1500	
QY	1527	TTGCTTTTGAAGAGAGAGCGCTTCTGAGGCTGTCATCTAGTGAGATACCTATGAGAA	1586	
DB	1501	TTGCTTTTGAAGAGAGAGCGCTTCTGAGGCTGTCATCTAGTGAGATACCTATGAGAA	1560	
QY	1587	AGGAGCTCGAGAGGGCGGAAATAGTGTAGTTGATTAATCTCATCTAGAGAGGAGCTTTC	1646	
DB	1561	AGGAGCTCGAGAGGGCGGAAATAGTGTAGTTGATTAATCTCATCTAGAGAGGAGCTTTC	1620	
QY	1647	CCATTTCTGCAGACCAAGCTTGTAGTAAACAACTGATAGATTATATTTTGGAGGTCAT	1706	
DB	1621	CCATTTCTGCAGACCAAGCTTGTAGTAAACAACTGATAGATTATATTTTGGAGGTCAT	1680	
QY	1707	CCAGTGACCTGGAAACAAAGTTCTGATCCAGAGGTGAGGATTCGGATGAGGAACTCAGG	1766	
DB	1681	CCAGTGACCTGGAAACAAAGTTCTGATCCAGAGGTGAGGATTCGGATGAGGAACTCAGG	1740	

Qy	1767	ATGATGTTTGTGATGATGTCTCAGTCTGACAGCTCAGACCTTGAACAACAAGACCCCTGAAG	1826
Db	1741	ATGATGTTTGTGATGATGATGTCTCAGTCTGACAGCTCAGACCTTGAACAACAAGACCCCTGAAG	1800
Qy	1827	GGCTTCACCTTTGGAACTCTTCTGCGAGTGTAGATCCCTTAATATCCCCAGCACTTTTACAG	1886
Db	1801	GGCTTCACCTTTGGAACTCTTCTGCGAGTGTAGATCCCTTAATATCCCCAGCACTTTTACAG	1860
Qy	1887	CAACAATTCAGACTGTGCCAGAAATGTTCCTGAAGAGCCCTTCGTATTCAGAGAAGGATT	1946
Db	1861	CAACAATTCAGACTGTGCCAGAAATGTTCCTGAAGAGCCCTTCGTATTCAGAGAAGGATT	1920
Qy	1947	TGTCTGCCAAGTCTGATCTAGAAATTCCTCCAGTCTCGAAGCCTTCCTGAGACCCCTG	2006
Db	1921	TGTCTGCCAAGTCTGATCTAGAAATTCCTCCAGTCTCGAAGCCTTCCTGAGACCCCTG	1980
Qy	2007	AGCATAGTCTCTGGGAGCAAGTGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCA	2066
Db	1981	AGCATAGTCTCTGGGAGCAAGTGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCA	2040
Qy	2067	AACGTGGAACTCATTCCTGTAATTCGTATGACCCCTACAAACCCCTTTAAATTTTAAAGCTC	2126
Db	2041	AACGTGGAACTCATTCCTGTAATTCGTATGACCCCTACAAACCCCTTTAAATTTTAAAGCTC	2100
Qy	2127	CTTTTCAAAATCAGGGGAAATGAGAAAGGCTGTCTGTGACTCAAGACCCCATCTCTGAGT	2186
Db	2101	CTTTTCAAAATCAGGGGAAATGAGAAAGGCTGTCTGTGACTCAAGACCCCATCTCTGAGT	2160
Qy	2187	CCATTGTGGCCATTTCTGAGTGTGCACACCTTACTTTCTCTGTAAGTGCAGCTGTGGGGA	2246
Db	2161	CCATTGTGGCCATTTCTGAGTGTGCACACCTTACTTTCTCTGTAAGTGCAGCTGTGGGGA	2220
Qy	2247	GCCAAAGAAAGTGAATGTCCAGACTCGGTGACGGTGACGTTCTTTCTGGAGGAGAACACA	2306
Db	2221	GCCAAAGAAAGTGAATGTCCAGACTCGGTGACGGTGACGTTCTTTCTGGAGGAGAACACA	2280
Qy	2307	CACATGTCAAAAGAAAAGGTAACTTCTTGNAGAGAGTTACTGAGTATTTATATAGTG	2366
Db	2281	CACATGTCAAAAGAAAAGGTAACTTCTTGNAGAGAGTTACTGAGTATTTATATAGTG	2340
Qy	2367	GTGATGAGGATCGCAAAGGACCATGGGAAGAAATTTGCAAGGGATGGAATGCAGGTTCCAGA	2426
Db	2341	GTGATGAGGATCGCAAAGGACCATGGGAAGAAATTTGCAAGGGATGGAATGCAGGTTCCAGA	2400
Qy	2427	AACGAATTCAGAAAACAGAAAGATGCTAATTCGGATATTCCTTGAACTTTGAAACAGAGAAA	2486
Db	2401	AACGAATTCAGAAAACAGAAAGATGCTAATTCGGATATTCCTTGAACTTTGAAACAGAGAAA	2460
Qy	2487	GAATGTTTTAATAGACTCCAGGGAACATGCTTCAAAGGACCTTAATGTTCTCTCAAGCAATGTT	2546
Db	2461	GAATGTTTTAATAGACTCCAGGGAACATGCTTCAAAGGACCTTAATGTTCTCTCAAGCAATGTT	2520
Qy	2547	GAGTTGCGACGCTGTAGTCTCTAGCTAGCATACACTCTCTTACCTGAGAGGTGTCTTTT	2606
Db	2521	GAGTTGCGACGCTGTAGTCTCTAGCTAGCATACACTCTCTTACCTGAGAGGTGTCTTTT	2580
Qy	2607	AAAAACAAATCTTGGCAGCTGTCTTTTGAACATTTTTTTTTTTTTTGAAGAAATGTAACTTGG	2666
Db	2581	AAAAACAAATCTTGGCAGCTGTCTTTTGAACATTTTTTTTTTTTTTTTTTGAAGAAATGTAACTTGG	2640
Qy	2667	ATCTAGTTTTAATTTTTTTTTTTTTTGGCAACATATCCACACTCAGAAACATTCAGGTTTGAAGCC	2726
Db	2641	ATCTAGTTTTAATTTTTTTTTTTTTTGGCAACATATCCACACTCAGAAACATTCAGGTTTGAAGCC	2700
Qy	2727	AGCCCTGATAATGAAGGATGAACATAGTGTGATTTCTAATCCTCCCTTTTTTGTATTTAGTT	2786
Db	2701	AGCCCTGATAATGAAGGATGAACATAGTGTGATTTCTAATCCTCCCTTTTTTGTATTTAGTT	2760
Qy	2787	GGATGTGCTTTTAAATGTCTTTTGCCTGCATGAGGTGGAAAGGGACCTTTTTTGAGTTGT	2846
Db	2761	GGATGTGCTTTTAAATGTCTTTTGCCTGCATGAGGTGGAAAGGGACCTTTTTTGAGTTGT	2820

Query Match	79.2%;	Score 2329;	DB 9;	Length 167079;
Best Local Similarity	99.8%;	Pred. No. 0;		

Qy	2847	CATTTTGCACCTTCAAAACCTATTTCTTGGAAAAACAATATTTATAGGGCTTTAAAGCCCA	2896
Db	2821	CATTTTGCACCTTTCACAAACCTATTTCTTGGAAAAACAATATTTATAGGGCTTTAAAGCCCA	2880
Qy	2907	TTTTCATTCTTAATCTAAATTAATGTGTGCTATCTG	2942
Db	2881	TTTTCATTCTTAATCTAAATTAATGTGTGCTATCTG	2916
RESULT 5			
AL606489/c			
LOCUS	AL606489	167079 bp	DNA linear PRI 08-NOV-2002
DEFINITION	Human DNA sequence from clone Rp11-739N20 on chromosome 1, complete sequence.		
ACCESSION	AL606489		
VERSION	AL606489.26	GI:24940211	
KEYWORDS	HTG.		
SOURCE	AL606489.26	GI:24940211	
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Pearce, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-NOV-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	On Nov 12, 2002 this sequence version replaced gi:2204251. ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk		

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

on the WormPeP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

RP11-799N20 is from the library RPO1-11.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

```

1. 167079
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-739N20"
/clone_lib="RP11-739N20"

```

ORIGIN

		Matches 2332; Conservative 0; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGCCTTACCGAGTCCCTTCGGGTATCGCGTTG	60		
Db	63775	ATTTTGGGCTTCGCTTCCACCGCACCGCGCGCTTACCGAGTCCCTTCGGGTATCGCGTTG	63716		
Qy	61	CTCAGGGGCTTTTCAACCCCTCTGTGTCAGTTCGGAACCACTCGCGAGGCGCTGGGGGACT	120		
Db	63715	CTCAGGGGCTTTTCAACCCCTCTGTGTCAGTTCGGAACCACTCGCGAGGCGCTGGGGGACT	63656		
Qy	121	CCTATCCATGTGTTGAAGCGTCGAGCCGACCTAGGGAACTCTCTCCCGCGCAGGATGGA	180		
Db	63655	CCTATCCATGTGTTGAAGCGTCGAGCCGACCTAGGGAACTCTCTCCCGCGCAGGATGGA	63596		
Qy	181	AGTCGATCAGTCGCGCGCTATTGCGCGGGCTGTTCTTCCTCTGTGTTCTGCGCCCGCTG	240		
Db	63595	AGTCGATCAGTCGCGCGCTATTGCGCGGGCTGTTCTTCCTCTGTGTTCTGCGCCCGCTG	63536		
Qy	241	CGCGATTTCGCTGCGCTCTGTGCTTTTCTGTGGCTCGAAGATCGGCCTGGAGCAGCGAC	300		
Db	63535	CGCGATTTCGCTGCGCTCTGTGCTTTTCTGTGGCTCGAAGATCGGCCTGGAGCAGCGAC	63476		
Qy	301	GCCACCGCTGGGCAAGGCGCGAGACTCTGTAGGCTTCCTCGAACTCCCGTCAACCTCCAGC	360		
Db	63475	GCCACCGCTGGGCAAGGCGCGAGACTCTGTAGGCTTCCTCGAACTCCCGTCAACCTCCAGC	63416		
Qy	361	CGCTGAGCGCGCGGCCCTTACTGTAGAGACTGTCAAGAAAGAGATGGAGCGGGGAC	420		
Db	63415	CGCTGAGCGCGCGGCCCTTACTGTAGAGACTGTCAAGAAAGAGATGGAGCGGGGAC	63356		
Qy	421	AGGCGGATCGCGGAAACGGCTTGGGCCCTCGGCGGGCTTCCGGTTCCTGGCCACCCCTTTT	480		
Db	63355	AGGCGGATCGCGGAAACGGCTTGGGCCCTCGGCGGGCTTCCGGTTCCTGGCCACCCCTTTT	63296		
Qy	481	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCGACGCTCTTGGCCCGGAAAATC	540		
Db	63295	CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCGACGCTCTTGGCCCGGAAAATC	63236		
Qy	541	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGCGAGACTCGGGTCAGTCTACTTGACGAA	600		
Db	63235	CGGGAAACCCACACTGCTTCTCTGCGCCAGCCGCGAGACTCGGGTCAGTCTACTTGACGAA	63176		
Qy	601	ACTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTCCTTCAGAAAGTGTCTAAATTTGGAG	660		
Db	63175	ACTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTCCTTCAGAAAGTGTCTAAATTTGGAG	63116		
Qy	661	CCAACTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTGCCTGAGTCTACAGCGC	720		
Db	63115	CCAACTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTGCCTGAGTCTACAGCGC	63056		
Qy	721	CTGTAGAGCCCTGAAGGGACGGGAGAAACACAGCCGCCCGCAGCGCAGAAATCTTTGAG	780		
Db	63055	CTGTAGAGCCCTGAAGGGACGGGAGAAACACAGCCGCCCGCAGCGCAGAAATCTTTGAG	62996		
Qy	781	TTGCGTGCAGCTTCGACTCTCCTCAGACCCCTCGGTCAACAGTCCCTTCGATTCGCTAGAGGA	840		
Db	62995	TTGCGTGCAGCTTCGACTCTCCTCAGACCCCTCGGTCAACAGTCCCTTCGATTCGCTAGAGGA	62936		
Qy	841	GGGGATCACTGGGCAATACTGCCGCCCGACCTTAAATTTGGAGCTTAAGGCCGAGGGAG	900		
Db	62935	GGGGATCACTGGGCAATACTGCCGCCCGACCTTAAATTTGGAGCTTAAGGCCGAGGGAG	62876		
Qy	901	TGCTTTTGACCCCTGCAGCACAGGCTTTTCTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960		
Db	62875	TGCTTTTGACCCCTGCAGCACAGGCTTTTCTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT	62816		
Qy	961	GTTCGCCAGTACCTTCAATCCCGTCTGTACTCTAAACCGGGAACTTGGCTCTTTCGCCCTC	1020		
Db	62815	GTTCGCCAGTACCTTCAATCCCGTCTGTACTCTAAACCGGGAACTTGGCTCTTTCGCCCTC	62756		
Qy	1021	TGGGCTCTTAACATTTCAACGATAGACAAATTTCAAGTGTGGTATCCTATTTCGTGAACCC	1080		
Db	62755	TGGGCTCTTAACATTTCAACGATAGACAAATTTTCAGTGTGGTATCCTATTTCGTGAACCC	62696		

Qy	1081	TTCTACTCGAGCTGCTTTCTTAGGCTAGAAGTCAGCTATCAGAACAGTGTGATGGAATAG	1140		
Db	62695	TTCTACTCGAGCTGCTTTCTTAGGCTAGAAGTCAGCTATCAGAACAGTGTGATGGAATAG	62636		
Qy	1141	CGAGGTAGTCGGCTTCAGACACTAACCCAGAGAGAGCTGCTGAGAGAGGACCAATTG	1200		
Db	62635	CGAGGTAGTCGGCTTCAGACACTAACCCAGAGAGAGCTGCTGAGAGAGGACCAATTG	62576		
Qy	1201	TCATCCCCAGCGCTGAGTCGAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACCTCT	1260		
Db	62575	TCATCCCCAGCGCTGAGTCGAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACCTCT	62516		
Qy	1261	TTCTACGGAAGGCTACCCAGAAATTCACCATTTTCGATGAAACCGCTGGAAATTCCTTCA	1320		
Db	62515	TTCTACGGAAGGCTACCCAGAAATTCACCATTTTCGATGAAACCGCTGGAAATTCCTTCA	62456		
Qy	1321	ACAGGCTAAACAGGGGCAAGATTTTACCCACCCCTGACAGGATATGGCTACCAAGCCT	1380		
Db	62455	ACAGGCTAAACAGGGGCAAGATTTTACCCACCCCTGACAGGATATGGCTACCAAGCCT	62396		
Qy	1381	GGAGGAGAAACACAGCCTTCTCCGGATGGATCCAAAACACTGCAGAGATAACCCAAACA	1440		
Db	62395	GGAGGAGAAACACAGCCTTCTCCGGATGGATCCAAAACACTGCAGAGATAACCCAAACA	62336		
Qy	1441	GTTTGTTCCTGCTGCTGGAGACATTCCTTGAAACACCCAGGAATCCACTGAAGAAAAAT	1500		
Db	62335	GTTTGTTCCTGCTGCTGGAGACATTCCTTGAAACACCCAGGAATCCACTGAAGAAAAAT	62276		
Qy	1501	AGAAATTAATTAACATCAGAGGTTCCACTTGTGGAAGAGAGAGCCCTTCGAGGCTG	1560		
Db	62275	AGAAATTAATTAACATCAGAGGTTCCACTTGTGGAAGAGAGAGCCCTTCGAGGCTG	62216		
Qy	1561	TCCATCTAGTGAGATACCTATGGAAGAGAGCCCTGAGAGGCGCGAATAAGTGTAGTTGA	1620		
Db	62155	TCCATCTAGTGAGATACCTATGGAAGAGAGCCCTGAGAGGCGCGAATAAGTGTAGTTGA	62156		
Qy	1621	TTACTCATACCTAGAAGGTGACCTTCCCATTTTCGCAGACCCAGCTTGTGTAGTAAACAACT	1680		
Db	62155	TTACTCATACCTAGAAGGTGACCTTCCCATTTTCGCAGACCCAGCTTGTGTAGTAAACAACT	62096		
Qy	1681	GATGATATTAATTTTGGAGGTCGATCCAGTGACCTTGGAAACAGTTCCTGATCCAGAGG	1740		
Db	62095	GATGATATTAATTTTGGAGGTCGATCCAGTGACCTTGGAAACAGTTCCTGATCCAGAGG	62036		
Qy	1741	TGAGATTTGGATGAGGAAGCTGAGGATGATGGTTTTGTAGTAGTATAGCTCACCTGTGAGA	1800		
Db	62035	TGAGATTTGGATGAGGAAGCTGAGGATGATGGTTTTGTAGTAGTATAGCTCACCTGTGAGA	61976		
Qy	1801	CTCAGACCTTTGAACAAGACCCCTGAAGGGCTTACCTTTTGGAACTCTTTCTGCACTGTAGA	1860		
Db	61975	CTCAGACCTTTGAACAAGACCCCTGAAGGGCTTACCTTTTGGAACTCTTTCTGCACTGTAGA	61916		
Qy	1861	TCCTTATAATCCCGCAAGACTTTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTTCTGA	1920		
Db	61915	TCCTTATAATCCCGCAAGACTTTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTTCTGA	61856		
Qy	1921	AGAGCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATTCCTCCCA	1980		
Db	61855	AGAGCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATTCCTCCCA	61796		
Qy	1981	GTCTGGGAAGCCTTCTCAGACCCCTGAGCATAGTTCCTGGGGAGGAAGATGCTGGGAATC	2040		
Db	61795	GTCTGGGAAGCCTTCTCAGACCCCTGAGCATAGTTCCTGGGGAGGAAGATGCTGGGAATC	61736		
Qy	2041	TAGTGAGATGAAGCAGAGGCTCTCAACTGTGAACTCAATCTGTGTAATTCGTATGACCC	2100		
Db	61735	TAGTGAGATGAAGCAGAGGCTCTCAACTGTGAACTCAATCTGTGTAATTCGTATGACCC	61676		
Qy	2101	CTACAAACCCCTTAAATTTTAAAGGCTCCCTTTTCAAAACATCAGGGGAAAAATGAGAAAGGCTG	2160		
Db	61675	CTACAAACCCCTTAAATTTTAAAGGCTCCCTTTTCAAAACATCAGGGGAAAAATGAGAAAGGCTG	61616		

```

Qy 2161 TCGTACTCAAGACCCCATCTGAGTCCATTGTGGCCATTCTTGAGTGTACACCTTACT 2220
Db 61615 TCGTACTCAAGACCCCATCTGAGTCCATTGTGGCCATTCTTGAGTGTACACCTTACT 61556
Qy 2221 TTCTTTGAAGTGCAGCTGTGGGAGCCAAAGAGTGAATGTCCACACTCGGTACAGCG 2280
Db 61555 TTCTTTGAAGTGCAGCTGTGGGAGCCAAAGAGTGAATGTCCACACTCGGTACAGCG 61496
Qy 2281 TGAGTCTCTTTCTTGAGAGAGACACACACATGTCAAAAGAAAAAAGGTAACTTCTT 2337
Db 61495 TGAGTCTCTTTCTTGAGAGAGACACACACATGTCAAAAGAAAAAAGGTAACTTCTT 61439

RESULT 6
AL627202/c
LOCUS Homo sapiens chromosome 1 clone RP11-174P23, 2 unordered pieces.
DEFINITION AL627202
ACCESSION AL627202
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Reference
1 (bases 1 to 328863)
Clark, S.
Direct Submission
Submitted (21-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:20268846.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA174P23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 170058 bases at least Q40
Consensus quality: 171825 bases at least Q30
Consensus quality: 172641 bases at least Q20
Insert size: 328763; sum-of-contigs
Insert size: 172230; 12.6% error; agarose-fp
Quality coverage: 2.67x in Q20 bases; sum-of-contigs Quality
coverage: 5.15x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6943: contig of 6943 bp in length
* 6944 7043: gap of 100 bp
* 7044 328863: contig of 321820 bp in length.

FEATURES
source
1. 328863
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-174P23"
/clone_lib="RPCI-11.1"
1..6943
/misc_feature
/note="assembly_fragment:03428"
7044..328863
/misc_feature
/note="assembly_fragment:03558"

```

ORIGIN

```

Query Match 79.2%; Score 2329; DB 2; Length 328863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2332; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATTTTGGGCTTGCCTTCCACCGCACAGCGGGCTACCCAGTCTCTTCGGGTATCGGTTG 60
Db 225562 ATTTTGGGCTTGCCTTCCACCGCACAGCGGGCTACCCAGTCTCTTCGGGTATCGGTTG 225503
Qy 61 CTCAGGGGCTTTTCAACCCCTCTGTGAGTCGGAACCATTCGCCGAGGCGCTGGGGGACT 120
Db 225502 CTCAGGGGCTTTTCAACCCCTCTGTGAGTCGGAACCATTCGCCGAGGCGCTGGGGGACT 225443
Qy 121 CCTATCCATGGTGTGAAGCGTCGAGCCGACCTAGGGAACCTCTCTCCCGCCAGGATGGA 180
Db 225442 CCTATCCATGGTGTGAAGCGTCGAGCCGACCTAGGGAACCTCTCTCCCGCCAGGATGGA 225383
Qy 181 AGTCGATCAGTCGCGGCTATTCGCGGGCTGTCTTCCTGCTGTGTCTGCGGCCGCTG 240
Db 225382 AGTCGATCAGTCGCGGCTATTCGCGGGCTGTCTTCCTGCTGTGTCTGCGGCCGCTG 225323
Qy 241 CGCATTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTGGAGCAGGAC 300
Db 225322 CGCATTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTGGAGCAGGAC 225263
Qy 301 GCCACGCTGGGCAAGCGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCCGACTCCAGC 360
Db 225262 GCCACGCTGGGCAAGCGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCCGACTCCAGC 225203
Qy 361 CGCTAGCGCGCGGCTTACTCTGAGAGACTGTCAAGAAAAAGAGAGATGAGCGGGGAC 420
Db 225202 CGCTAGCGCGCGGCTTACTCTGAGAGACTGTCAAGAAAAAGAGAGATGAGCGGGGAC 225143
Qy 421 AGCGGATCGCGAAGCGGCTTGGCGCTCGGGGGGCTTCCGGTCTTGGCCACCTTTT 480
Db 225142 AGCGGATCGCGAAGCGGCTTGGCGCTCGGGGGGCTTCCGGTCTTGGCCACCTTTT 225083
Qy 481 CCCTCGCGGATCGCAAGCAGGCTTCTTAAGTTCCGAGCGCTCTTTGGCCCGGAAATC 540
Db 225082 CCCTCGCGGATCGCAAGCAGGCTTCTTAAGTTCCGAGCGCTCTTTGGCCCGGAAATC 225023
Qy 541 CGGGAACCCCACTCTCTTCTGCCAGCCCGAGACTCGGGTCACTTCTGAGCGAA 600
Db 225022 CGGGAACCCCACTCTCTTCTGCCAGCCCGAGACTCGGGTCACTTCTGAGCGAA 224963
Qy 601 ACTGCTCTCCAGCTCTTGGCGCTCCCGGCTTCTCAGAGGCTCTTAATTTGGAG 660
Db 224962 ACTGCTCTCCAGCTCTTGGCGCTCCCGGCTTCTCAGAGGCTCTTAATTTGGAG 224903
Qy 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTGTCTGAGTCTACAGCG 720
Db 224902 CCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTGTCTGAGTCTACAGCG 224843
Qy 721 CTGAGAGCCCTGAAGGAGCGGAGAAACAGCGCGCCCGCCACAGCGCAGAAATCTTTGAG 780
Db 224842 CTGAGAGCCCTGAAGGAGCGGAGAAACAGCGCGCCCGCCACAGCGCAGAAATCTTTGAG 224783
Qy 781 TTCGCTGAGCTGAGCTCTCTCAGACCCCTCGGTCACAGTCCCTTGTATTGGCTAGAGGA 840
Db 224782 TTCGCTGAGCTGAGCTCTCTCAGACCCCTCGGTCACAGTCCCTTGTATTGGCTAGAGGA 224723
Qy 841 GGGGATCCCACTGGCAATACTCGCCCCCGCCAGACCTTAAATTTGGAGCTTAAGCCAAAGGAAG 900
Db 224722 GGGGATCCCACTGGCAATACTCGCCCCCGCCAGACCTTAAATTTGGAGCTTAAGCCAAAGGAAG 224663
Qy 901 TGCTTTGGAGCCCTGAGAGCAGGCTTTTCTTTAGAGCAGAGCTGTGGGAGTGGAGCT 960
Db 224662 TGCTTTGGAGCCCTGAGAGCAGGCTTTTCTTTAGAGCAGAGCTGTGGGAGTGGAGCT 224603
Qy 961 GTTGGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGAGACTTGGCTCTTTCGCCCTC 1020
Db 224602 GTTGGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGAGACTTGGCTCTTTCGCCCTC 224543

```

Qy 1021 TGGGCTCTAAACATTTCAACGCATAGACAAATTTTCAGTGTGGTATCCTATTTCCTGGAACCC 1080
Db 224542 TGGGCTCTAAACATTTCAACGCATAGACAAATTTTCAGTGTGGTATCCTATTTCCTGGAACCC 224483
Qy 1081 TTCCTACCTGGAGCTGCTTTCTTAGCTAGAACTAGAGCTATCAGAACAGTGTGGAATAG 1140
Db 224482 TTCCTACCTGGAGCTGCTTTCTTAGCTAGAACTAGAGCTATCAGAACAGTGTGGAATAG 224423
Qy 1141 CGAGGTAGTGGCTTCAGACACTAAACCCAGAGAGAGCTGCTGAGAGAGGACCATG 1200
Db 224422 CGAGGTAGTGGCTTCAGACACTAAACCCAGAGAGAGCTGCTGAGAGAGGACCATG 224363
Qy 1201 TCATCCCGAGCGCTGAGTGCAGAACTCATTCGGGCTCGTGGCAGGAGATGTCACCTCT 1260
Db 224362 TCATCCCGAGCGCTGAGTGCAGAACTCATTCGGGCTCGTGGCAGGAGATGTCACCTCT 224303
Qy 1261 TTCTACGGAAGGCTTACCAGAAATTCACCATCTTCGCATGAACCGGCTGGAAATTCCTTCA 1320
Db 224302 TTCTACGGAAGGCTTACCAGAAATTCACCATCTTCGCATGAACCGGCTGGAAATTCCTTCA 224243
Qy 1321 ACAGCTTAACAGGGGCAAGATTTACCCACCCCTGACAGGATATGCTACCAAGCT 1380
Db 224242 ACAGCTTAACAGGGGCAAGATTTACCCACCCCTGACAGGATATGCTACCAAGCT 224183
Qy 1381 GGAGGAGGAACACAGCTTCTCCGATGGATCCAAACACTGCAGAGATAACCCAAACA 1440
Db 224182 GGAGGAGGAACACAGCTTCTCCGATGGATCCAAACACTGCAGAGATAACCCAAACA 224123
Qy 1441 GTTTGTTCTGCTGCTGGAGACATTCCTGGAAACCCAGGAATCCACTGAAGAAAAAT 1500
Db 224122 GTTTGTTCTGCTGCTGGAGACATTCCTGGAAACCCAGGAATCCACTGAAGAAAAAT 224063
Qy 1501 AGAATATTAACACAGAGTTCCACTTGTCTTTGGAAGAGAGCCCTTCTGAGGGCTG 1560
Db 224062 AGAATATTAACACAGAGTTCCACTTGTCTTTGGAAGAGAGCCCTTCTGAGGGCTG 224003
Qy 1561 TCCATCTAGTAGATACCTATGGAAGAGGCTGAGAGGGCCGAATAAGTGTAGTTGA 1620
Db 224002 TCCATCTAGTAGATACCTATGGAAGAGGCTGAGAGGGCCGAATAAGTGTAGTTGA 223943
Qy 1621 TTACTCATACCTAGAAAGTGCATTCCTCCAGACAGAGCTTGTAGTAAACAACT 1680
Db 223942 TTACTCATACCTAGAAAGTGCATTCCTCCAGACAGAGCTTGTAGTAAACAACT 223883
Qy 1681 GATAGATTATATTTTGGAGGTGCATCCAGTACCTGGAAACAGTTCTGATCCAGAGG 1740
Db 223882 GATAGATTATATTTTGGAGGTGCATCCAGTACCTGGAAACAGTTCTGATCCAGAGG 223823
Qy 1741 TGAGGATTTGGATGAGGAAGCTGAGGATGATGGTTTGTAGTAGTAGTCACTGTCA 1800
Db 223822 TGAGGATTTGGATGAGGAAGCTGAGGATGATGGTTTGTAGTAGTAGTCACTGTCA 223763
Qy 1801 CTCAGACCTTGAACAGACCCCTGAAGGCTTCACTTTGGAACCTTTCTGAGGTGAGA 1860
Db 223762 CTCAGACCTTGAACAGACCCCTGAAGGCTTCACTTTGGAACCTTTCTGAGGTGAGA 223703
Qy 1861 TCCTTATATCCCGAAGCTTTAGAGCAACATTCAGACTGCTGCCAGATTTGTCCTGA 1920
Db 223702 TCCTTATATCCCGAAGCTTTAGAGCAACATTCAGACTGCTGCCAGATTTGTCCTGA 223643
Qy 1921 AGAGCCTTCTGATTCAGAGAAGATTTGTTCGGCAAGCTGATCTGAGAGAAATTCCTCCA 1980
Db 223642 AGAGCCTTCTGATTCAGAGAAGATTTGTTCGGCAAGCTGATCTGAGAGAAATTCCTCCA 223583
Qy 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAGATGATGGGAATC 2040
Db 223582 GTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAGATGATGGGAATC 223523
Qy 2041 TAGTCAGATGAGCAGAGGCTCAAACTGTGGAACCTCATCTCTTAATCTGTAGTACCC 2100
Db 223522 TAGTCAGATGAGCAGAGGCTCAAACTGTGGAACCTCATCTCTTAATCTGTAGTACCC 223463

Qy 2101 CTACAAACCCCTTTAAATTTTAAAGCTCCTCTTTTCAAAACATCAGGGGAAAATGAGAAGGCTG 2160
Db 223462 CTACAAACCCCTTTAAATTTTAAAGCTCCTCTTTTCAAAACATCAGGGGAAAATGAGAAGGCTG 223403
Qy 2161 TCGTGACTCAAGAACCCTCTCTGAGTCCATGTGGCCCATTTCTGAGTGTGCACACTTACT 2220
Db 223402 TCGTGACTCAAGAACCCTCTCTGAGTCCATGTGGCCCATTTCTGAGTGTGCACACTTACT 223343
Qy 2221 TTCTTTGAAGTGCAGCTGTGGGAGCCCAAGAAAGTGAATGCTCCAGACTCGGTACAGCG 2280
Db 223342 TTCTTTGAAGTGCAGCTGTGGGAGCCCAAGAAAGTGAATGCTCCAGACTCGGTACAGCG 223283
Qy 2281 TGAGCTTCTTTCTCGAGGAAGACACACACATGTCAAAAGAAAAGGTAACTTCTTCT 2337
Db 223282 TGAGCTTCTTTCTCGAGGAAGACACACACATGTCAAAAGAAAAGGTAACTTCTTCT 223226

RESULT 7
LOCUS BC058078
DEFINITION Mus musculus protein phosphatase 1, regulatory (inhibitor) subunit
15b, mRNA (cdna clone MGC:66517 IMAGE:5830215), complete cds.
ACCESSION BC058078
VERSION BC058078.1 GI:34785997
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 5367)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.I.,
Scheetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S.,
Carninci,P., Prange,S., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.M., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalusi,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

TITLE Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL 22388257
MEDLINE 1479732
PUBMED
REFERENCE 2 (bases 1 to 5367)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NTH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbe-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (IULNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven

Nees, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Teai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAK Plate: 126 Row: n Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27597066.

FEATURES

source

Location/Qualifiers
1..5367
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66517 IMAGE:6830215"
/tissue_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NIH_BMAP_F00"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

gene

1..5367
/gene="Pppl15b"
/db_xref="LocusID:108954"
/db_xref="MGI:2444211"
365..1591

CDS

/codon_start=1
/product="Pppl15b protein"
/protein_id="AAH58078.1"
/db_xref="GI:34785998"
/db_xref="LocusID:108954"
/db_xref="MGI:2444211"
/translation="METHTRHRRKPGRLGSPRLPLRRSHACSEPPPPSSRQNP
GNSLPERRRYWTKLSQLALPSLFQKLLWSOLFGLIPTRWLDFAASYSALRA
LRGREAAPFVQSLSSLDSEDLVSSLDWLRGLQWQSSSDELEKLKXQERA
LDSAAFTPLLEQLWGVLLPSSLSGLSHRELDSSSGLSPVQSLGNFKVSYLLN
PSYDYLPLQGLRQCSAGGQFVGFRTLPESCYLSEDCHPQLPRAEMSATAWRR
PPLSTGLPELHLRMRLEPLQANKQELPTPDQNGYHSLSEHNLRLMDPOHCTD
NPAQAVSPADRPPTPKKPELVQVSPQSSLSFCELPVEKECEDHTNATDLS
RGESLPVSTRPVATN"

ORIGIN

Query Match 46.4%; Score 1365.6; DB 10; Length 5367;
Best Local Similarity 73.0%; Pred. No. 0;
Matches 1975; Conservative 0; Mismatches 649; Indels 80; Gaps 14;
Qy 96 CCATCGCGAGCGCGTGGGGGACCTCTATCCATGGTGTGTAAGCGTCGAGCCACTAGG 155
Dy 54 CCGTCGCGCGCGCGAGGAGGGTCTTCTATATGGTGGAGCGATCTCACACGCGCTAGG 113
Qy 156 GAACCTCTCTCC-CGCGAGATGGAAGTGCATCAGTCGCGCTATTGCGGGGTGT 214
Dy 114 ACGTCTCTCTCCCTAGCCGGATGGAACCTAACCGGTCGACCGCTTGGCGGGGCTC 173
Qy 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGGCAATCGCTGCCCTCTGTGCTTTCTGCTGG 274
Dy 174 TGGGCGGTCCGCTGCAGCACTCGTTGCGGAAGCGCGCTCTCTGGGCTCTCTGCGG 233
Qy 275 CTCGAAGATCGGCTTGAGAGAGCAGCCACCGCTGGGCAAGCGCGAGACTCTGTAGCT 334
Dy 234 CGCGGAATCGACTGCACTGCCACCTCCGTGGCTGGGCAAGCGGAGACTGTGTAGACC 293
Qy 335 TCCTCCGAATCCGTCGACCTCCAGCGCTGAGCGCGCGCGCTTACTCTGAGAGACTGTC 394
Dy 294 TCGGATCCAGCTCGGCTGACCGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAGCGGCC 353
Qy 395 AAGAAAAGGAGATGAGCGCGGAGACAGCGGATCGCGAAACGCGCTTGGCCCTCGGCG 454
Dy 354 AAG-GAAGGAGATGAGACAGGACAGGACAGGCGCGGAGCGCTTGGCCCTCGGCTG 412
Qy 455 GCCTTCGGTTCGCGCCACCTTTTTCCTCGGCGATCGAAGCGGCTCTTCTAAGTTC 514

Db 413 GGCTCTCGTGTTCGGCTGCCCC---TTCTTTCGGGATCGACGCCCTGCTCTTCGGAGTTC 469
Qy 515 CCAGACGCTCTTGTGGCCCGGAAACTCCGGGAACCCACACTGCTTCTCTGCCCCAGCCC 574
Db 470 CCGCGGCTTCTCTCTCGACAAATCCGGGAA-----CTCCGCTCTGCC 514
Qy 575 GAGACTCGGCTAGTTACTTGGAGGAAATGCTCTCCAGCTCTTGGCGGCTCCCCGGA 634
Db 515 GAGCGTGGACCAAGGTAAGTTCGACCAATGCTTCTCAGCTCTTCCCTGCTCCCTAGC 574
Qy 635 TTCTTTCAGAGGTGCTTAATTTGGAGCAACTTTTGGGTGGAATGTTTCCGACCAAGTGG 694
Db 575 CTATTCAGAGAGTGTCTGCTTTGGAGCCAGCTTTTGGGGGCTGATTTCTACCAAGTGG 634
Qy 695 CTAGATTTTGTCTGAGTCTACAGCGCTTGGAGGAGTCCCTGAGGAGGAGGAGGAGTTC 754
Db 635 CTAGATTTTGTCCGCAAGTTACAGCGCTTGGAGGAGTTCGAGAGGAGGAGGAGTTCG 694
Qy 755 GCGCCACAGCGCAGAAATCTTTGAGTTCGCTGAGTCTGAGTCTCAGACCCCTCGGTTC 814
Db 695 GCTCCACAGCGTGCAGAAAGTCTCTGAGTTCAGTCCGCTGAGTCTTTCGGAAGACTTGG 754
Qy 815 ACCAGTCCCTTCTGATTTGGCTAGAGGAGGATCCACTGGCAATACTCGCCGCCAGACCTA 874
Db 755 GTAGTCTCTTGTATTTGGCTAGAGGAGGAGTCCAGTGGAGTGTCTGCTCTCAGACCTG 814
Qy 875 AAATTTGAGCTTAAGGCCAAGGAGTGTCTTGGACCTTGCAGCAGAGGCTTTTCTCTTA 934
Db 815 GAGTAAACTCAAGGCCCGCAGGAAAGAGCTTTAGACTCTGCGCGGCCACTTTTCTCTG 874
Qy 935 GAGCAGAGCTGTGGGAGTGGAGCTGTTCGCCAGTAGCTTCAATCCGCTGTACTCT 994
Db 875 GAGCAGCAGCTGTGGGAGTGGAGTGTCTGCCCCAGTAGCTTCAAGCTGTGTCTAGTCTCC 934
Qy 995 AACCGGGAATCTGCTCTTCCGCTCTGCGCTCTGAGCTCTTAAACATTCACGCAATAGCA 1054
Db 935 CACCGAGACTTGAATCTTCTCTCTGCGCTCTGAGCTCTGAGCTTCAAGCTTGTAGTCTTC 994
Qy 1055 AGTGTGTATCTTATTTGCTGAAACCTTCTTACCTGAGCTGTCTTCTAGGCTAGAAATC 1114
Db 995 AAGTAGTTTCTCTATCTCTGAAACCTTCTCTACCTGAGCTTACCTTCCCGAGTGTAGGCTG 1054
Qy 1115 AGCTATCAGAACTAGTGAATAGGAGTGTGCTTCCAGACACTTAACCCAGAG 1174
Db 1055 CGCTGTGAGAGCAGCTGGAGTGGCCAGTTTGTGGTTCGGAACACTTAACCCAGAG 1114
Qy 1175 AGCAGCTGCTGAGAGAGGACATTTGTCATCCCGCTGAGTGTGAGAACTCATTTCCG 1234
Db 1115 AGCTGTATCTTTCTGAAGATGTTGTACCTTACCTTACCGCTTGGCGGAGAGATGTGGCA 1174
Qy 1235 GCCTCGTGGCAGGATGTCCACCTCTTTCTAAGGAGGCTTACAGAAATTCACCACTTT 1294
Db 1175 ACCGCTTGGAGAGTGTCCGCTCTCTCTACAGAGGCTGCGGAAATCCACACCTT 1234
Qy 1295 CGATGAACCGCTGGAATCTTCAACAGGCTTAAAGGGGCAAGATTTACCCACCCCT 1354
Db 1235 CGTATGAACCGCTGGAATTTCTT---CCAGGCTTAAAGAGGCAAGATTTACCCACCCCT 1291
Qy 1355 GAGCAGGATTAAGTGTACCACTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1414
Db 1292 GACCAAGATTAAGTGTATCTATGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1351
Qy 1415 AAACACTGAGAGATAACCCCAACACAGTGTGTTGTTCTGCTGTGAGAGCAATTTCTGGAAC 1474
Db 1352 CAACATTTGCACAGATAAACCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1404
Qy 1475 ACCAGGATTCAGTGAAGAAATAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1534
Db 1405 -----GGAGCCCACTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1451
Qy 1535 GAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1594
Db 1452 -----AGAGCCCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1504

1595 GGAGAGGCGGAATAAGTGTAGTTGATTACTCATACCTAGAGGTGACCTTCCCAATTCT 1654
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1505 GAAGAGGACCACTAATAGCACTGACCTCTCAGATAGAGGAGAGGCTTCTGTTTCT 1564
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1655 GCCAGACAGCTTGTAGTAAACAACTGATAGATTATATTTTGGGAGGTGCATCCAGTGAC 1714
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1565 ACCAGACAG -TTGTAGCAACAACTGATAGATTATATTTTGGGAGGCGCCCAAGTGAC 1623
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1715 CTGGAACCAAGTCTGTATCCAGAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGAGT 1774
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1624 TTGGAAGCCAGCTCTGTATCTGAAAGTGAAGATTTGGGGCGAGGAACCTTGAGGACGATGGC 1693
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1775 TTGTATAGTATGATGCTCACTGTGCAGACTCAGACCTTGAACAGAGACCTTGAAGGGCTTCAC 1834
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1684 TTGTATAGGATGGCTCTCTGTCTGTATCAGACGTGGAAACAGGACTCGAAGGCCCTTCAC 1743
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1835 CTTTGGAACTCTTTCTGCAGTGTAGATCCTTATATATCCCAAGAACTTTTACAGCAACAAT 1894
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1744 CTTTGGAACTCTTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTTACAGCCAGAT 1803
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1895 CAGACTGCTGCCAGAAATTTCTTCCAGTCTGGAAGCCTTCTGTGAGACCCCTTGAGCATAGT 1954
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1804 CAGACGGCTGCCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGAGCATCTCTGGTCTGGC 1863
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1955 AAGTCTGTATCAGAGAAATTTCTTCCAGTCTGGAAGCCTTCTGTGAGACCCCTTGAGCATAGT 2014
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1864 AGCTGTGTAGGGA---GCTGTGAGGAGGACCCCTTCCGGAGACCCCGACCATAGT 1920
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2015 TCTGGGAGGAGATGACTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGG 2074
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1921 TCCGGGAGGAGATGACTGGGAACCGAGTGCAGATGAAGCAGAGAACTTTAAATTTGTGG 1980
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2075 AACTCATTTCTGAATCTGATGACCCCTTACACCCCTTTAAATTTTAAAGCTCTTTTCAA 2134
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1981 AACTCTTTCTGTCTGATCTGAGACCCCTTACACCCCTTTAAATTTTAAAGCTCTTTTCAA 2040
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2135 ACATCAGGGGAAATGAGAAAGGCTGTCTGTACTCAAGACCCCATCTGAGTCCATTTGTG 2194
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2041 CGTTCAGGAGAAATTTGNAAGCCGTGAGACTCAAGGCTCTTCTGAGGCTACAGTG 2100
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2195 GCCATTTCTGAGTGTACACCTTACTTTCTGTAAGGTGAGCTGTTGGGAGCCCAAGAA 2254
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2101 GCCTTCTCTGGCCATCATACCTTACTTCTTGAAGGCCAGCTGTTAGAGAGCCCAAGAA 2160
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2255 AGTGAATGTCAGATCCGTACAGGTGACGTTCTTTCTGAGGAGAGACACACATGTC 2314
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2161 GATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTCTTCTGGAAGAAGATACACCCATATC 2220
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2315 AAAAGAAAAAGGTAACCTTCTTGAAGAAGTTACTGAGTATTTATTAAGTGGTGATGAG 2374
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2221 AAGAGAAAAAGTAACTTCTTGAAGAAGTTACTGAGTATTTATTAAGTGGTGATGAG 2280
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2375 GATCGAAAGGACCAATGGGAAGATTTGCAAGGGATGAGTGCAGGTTCCAGAACCAAT 2434
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2281 GATCGAAAGGACCAATGGGAAGATTTGCAAGGGATGAGTGCAGGTTCCAGAACCAAT 2340
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2435 CAAAGAAACAGAGATGCTATTGGATTTGCTTGACATTTGACACAGAGAAAGATGTTT 2494
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2341 CAAAGAAACAGAGATGCTATTGGATTTGCTTGACATTTGACACAGAGAAAGATGTTT 2400
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2495 AATAGACTCCAGGGAACATGCTTCAAGGACTTAATGTTTCTCAAGCAATGTTGAGTTGGC 2554
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2401 AATAGACT---GAGGATCGAGTCAAGGACTTACTGTTGTACACCAATGTTAAGAGTG 2456
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2555 AGCCTGTAGTCTAGTACATACCTCTTACCTGAGAGGTTGCTTTTAAACAA 2614
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2457 A-ACAGCCTGCAACCCGTCGCCACTCTGTCTTCTTCTGAGA-GTTTCCCTTTAAACAA 2514
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2615 ATCTTGGCAGCTGCTCTTTGACATTTTCTTTTATAGAGAAATGTAATCTGATCTAGT 2674
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2515 ACATGGGACGCTGCTCTGGACATGTTTAAAGAAACAACTTGTATCTAGAGATGCAGT 2574

2675 TAATTTTCTTTTTCACATATCCCACTCAGAAACATTCAGGTTTGAAGCCACCTGA 2734
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2575 TTGATTATTTTGGTAAATGTTCTCATTAGAA-----ACACCAACTCCGA 2620
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2735 TAATGAAGGAGAACTAGTGTGATTCTTAATCTCCCTTTTTCATTTAGTTAGTGTGATGC 2794
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2621 TAATGAAGAAT-----CTCTTATCTGTAATCTCTCTCTTTTCCCTATTAGTTAGTGTGG 2674
 Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 2795 TTTT 2798
 Db | | | |
 2675 GTTT 2678

AC048369 166233 bp DNA linear HTG 25-JUN-2000
 Homo sapiens chromosome 1 clone RP11-284G5 map 1, WORKING DRAFT
 AC048369
 AC048369.2 GI:8705131
 HTG: HTGS PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 1, clone RP11-284G5
 Unpublished
 2 (bases 1 to 166233)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2000 this sequence version replaced gi:7549699.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L8045
 Center clone name: 284_G_5
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151910 bases at least Q40
 Consensus quality: 158917 bases at least Q30
 Consensus quality: 161704 bases at least Q20

TITLE
 JOURNAL
 COMMENT

Insert size: 178000; agarose-fp
Insert size: 163433; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1069: contig of 1069 bp in length
* 1070 1169: gap of 100 bp
* 1170 2738: contig of 1569 bp in length
* 2739 2838: gap of 100 bp
* 2839 4143: contig of 1305 bp in length
* 4144 4243: gap of 100 bp
* 4244 5489: contig of 1246 bp in length
* 5490 5589: gap of 100 bp
* 5590 6873: contig of 1284 bp in length
* 6874 6973: gap of 100 bp
* 6974 9153: contig of 2180 bp in length
* 9154 9253: gap of 100 bp
* 9254 10555: contig of 1302 bp in length
* 10556 10655: gap of 100 bp
* 10656 13995: contig of 3340 bp in length
* 13996 14095: gap of 100 bp
* 14096 16689: contig of 2594 bp in length
* 16690 16789: gap of 100 bp
* 16790 19295: contig of 2506 bp in length
* 19296 19395: gap of 100 bp
* 19396 22797: contig of 3402 bp in length
* 22798 22897: gap of 100 bp
* 22898 26454: contig of 3556 bp in length
* 26455 31369: contig of 4816 bp in length
* 31370 31469: gap of 100 bp
* 31470 35528: contig of 4059 bp in length
* 35529 35629: gap of 100 bp
* 35630 40532: contig of 4903 bp in length
* 40533 40632: gap of 100 bp
* 40633 45154: contig of 4523 bp in length
* 45155 45254: gap of 100 bp
* 45255 50400: contig of 5146 bp in length
* 50401 50500: gap of 100 bp
* 50501 56357: contig of 5857 bp in length
* 56358 56457: gap of 100 bp
* 56458 62667: contig of 6210 bp in length
* 62668 62768: gap of 100 bp
* 62769 69282: contig of 6515 bp in length
* 69283 76270: gap of 100 bp
* 76271 76370: gap of 100 bp
* 76371 84430: contig of 8060 bp in length
* 84431 84531: gap of 100 bp
* 84532 92060: contig of 7530 bp in length
* 92061 92160: gap of 100 bp
* 92161 101111: contig of 8951 bp in length
* 101112 101212: gap of 100 bp
* 101213 109009: contig of 7798 bp in length
* 109010 109109: gap of 100 bp
* 109110 120552: contig of 11443 bp in length
* 120553 120652: gap of 100 bp
* 120653 130699: contig of 10047 bp in length
* 130700 130799: gap of 100 bp
* 130800 145936: contig of 15137 bp in length
* 145937 146037: gap of 100 bp
* 146038 166233: contig of 20197 bp in length.

FEATURES
source
1. .166233
/organism="Homo sapiens"

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1"
/clone="RP11-284G5"
/clone_lib="RPC1-11 Human Male BAC"
1. .1069
/note="assembly_fragment"
1170. .2738
/note="assembly_fragment"
2839. .4143
/note="assembly_fragment"
4244. .5489
/note="assembly_fragment"
5590. .6873
/note="assembly_fragment"
6974. .9153
/note="assembly_fragment"
9254. .10555
/note="assembly_fragment"
clone_end:17
vector_side:right"
10656. .13995
/note="assembly_fragment"
14096. .16689
/note="assembly_fragment"
16790. .19295
/note="assembly_fragment"
19396. .22797
/note="assembly_fragment"
22898. .26453
/note="assembly_fragment"
26554. .31369
/note="assembly_fragment"
31470. .35528
/note="assembly_fragment"
35629. .40531
/note="assembly_fragment"
40632. .45154
/note="assembly_fragment"
45255. .50400
/note="assembly_fragment"
50501. .56357
/note="assembly_fragment"
56458. .62667
/note="assembly_fragment"
62768. .69282
/note="assembly_fragment"
69383. .76270
/note="assembly_fragment"
76371. .84430
/note="assembly_fragment"
84531. .92060
/note="assembly_fragment"
92161. .101111
/note="assembly_fragment"
101212. .109009
/note="assembly_fragment"
109110. .120552
/note="assembly_fragment"
120653. .130699
/note="assembly_fragment"
130800. .145936
/note="assembly_fragment"
146037. .166233
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

ORIGIN

Query Match 44.3%; Score 1303.4; DB 2; Length 166233;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	694	GCTAGATTTTGTGGAGTCTACAGCGCCTGAGAGCCCTGAAGGACCGGAGAAACACAGC	753
Db	2839	GCTAGATTTTGTGGAGTCTACAGCGCCTGAGAGCCCTGAAGGACCGGAGAAACACAGC	2898
Qy	754	CGCCCCCAGCGCAGAAATCTTTGAGTTCGCTGAGCTCGACTCCTCAGACCCCTCGGT	813
Db	2899	CGCCCCCAGCGCAGAAATCTTTGAGTTCGCTGAGCTCGACTCCTCAGACCCCTCGGT	2958
Qy	814	CACAGTCCCTTGATTCGCTAGAGGAGGGGATCCACTGGCAATACCTCGCCCCAGAGCT	873
Db	2959	CACAGTCCCTTGATTCGCTAGAGGAGGGGATCCACTGGCAATACCTCGCCCCAGAGCT	3018
Qy	874	AAAAATTGAGAGCTTAAGGCGCAAGGAGAGTCTTTGGACCCCTGAGCAGCAGGCTTTCTCTT	933
Db	3019	AAAAATTGAGAGCTTAAGGCGCAAGGAGAGTCTTTGGACCCCTGAGCAGCAGGCTTTCTCTT	3078
Qy	934	AGAGCAGCAGCTGTGGGAGTGGAGCTGTTCGCCAGTAGCCCTTCAATCCCGTCTGTATCTC	993
Db	3079	AGAGCAGCAGCTGTGGGAGTGGAGCTGTTCGCCAGTAGCCCTTCAATCCCGTCTGTATCTC	3138
Qy	994	TAAACGGGAACCTTGGCTCTTCCCTCTGGGCTCTAAACATTTCAACGCATAGACAATTT	1053
Db	3139	TAAACGGGAACCTTGGCTCTTCCCTCTGGGCTCTAAACATTTCAACGCATAGACAATTT	3198
Qy	1054	CAGTGTGGTATCTTATTTGCTGAACCTTCTTACCTGGAGCTCTTCTTAGGCTAGAACT	1113
Db	3199	CAGTGTGGTATCTTATTTGCTGAACCTTCTTACCTGGAGCTCTTCTTAGGCTAGAACT	3258
Qy	1114	CAGCTATCAGAAACAGTGATGGAATATGCGAGTGTGCGGCTTCCAGACACTAAACCCAGA	1173
Db	3259	CAGCTATCAGAAACAGTGATGGAATATGCGAGTGTGCGGCTTCCAGACACTAAACCCAGA	3318
Qy	1174	GAGCAGCTCCCTGAGAGGACCATTTGCTATCCCGGCTGAGTGGAGCAACTCATTC	1233
Db	3319	GAGCAGCTCCCTGAGAGGACCATTTGCTATCCCGGCTGAGTGGAGCAACTCATTC	3378
Qy	1234	GGCTCTGTGGAGGATGTCACCTCTTTCTACGGAAGCCCTTACAGAAATTCACCATCT	1293
Db	3379	GGCTCTGTGGAGGATGTCACCTCTTTCTACGGAAGCCCTTACAGAAATTCACCATCT	3438
Qy	1294	TGCGATGAACCGGCTGGAATTCCTTCAACAGGCTAACAGGGGCAAGATTTACCCACCC	1353
Db	3439	TGCGATGAACCGGCTGGAATTCCTTCAACAGGCTAACAGGGGCAAGATTTACCCACCC	3498
Qy	1354	TGACAGGATATGCTACACAGCTTGAGGAGGAACACAGCTTCTCCGATGGATCC	1413
Db	3499	TGACAGGATATGCTACACAGCTTGAGGAGGAACACAGCTTCTCCGATGGATCC	3558
Qy	1414	AAAACACTGCAGAGATAACCCACACAGTTCCTGCTGCTGGAGACATTCCTGGAAA	1473
Db	3559	AAAACACTGCAGAGATAACCCACACAGTTCCTGCTGCTGGAGACATTCCTGGAAA	3618
Qy	1474	CACCCAGGAATCCACTGAAGAAAAATAGAATTAATTAACAGAGGTTCCACTTGCCTT	1533
Db	3619	CACCCAGGAATCCACTGAAGAAAAATAGAATTAATTAACAGAGGTTCCACTTGCCTT	3678
Qy	1534	GGAGAGAGAGCCCTTCTGAGGGCTGTCCTAGTAGGATACCTATGGAAAAGGAGCC	1593
Db	3679	GGAGAGAGAGCCCTTCTGAGGGCTGTCCTAGTAGGATACCTATGGAAAAGGAGCC	3738
Qy	1594	TGGAGAGGGCCGAATAAGTGTAGTTGATTTACTCATACCTAGAAAGTGACCTTCCCATTC	1653
Db	3739	TGGAGAGGGCCGAATAAGTGTAGTTGATTTACTCATACCTAGAAAGTGACCTTCCCATTC	3798
Qy	1654	TGCCAGACAGCTTGTAGTAAACAACTGATAGATTATATTTGGAGGTGCATCCAGTGA	1713
Db	3799	TGCCAGACAGCTTGTAGTAAACAACTGATAGATTATATTTGGAGGTGCATCCAGTGA	3858
Qy	1714	CTGTGAAAACAAGTCTGTATCCAGAGGTGAGGATTTGGATGAGGAGCTGAGGATGATGG	1773
Db	3859	CCTGGAAAACAAGTCTGTATCCAGAGGTGAGGATTTGGATGAGGAGCTGAGGATGATGG	3918
Qy	1774	TTTTGATAGTATAGTCTCACTGTGAGACTCAGACCTTGAAACAAGACCTCGAAGGGCTTCA	1833
Db	3919	TTTTGATAGTATAGTCTCACTGTGAGACTCAGACCTTGAAACAAGACCTCGAAGGGCTTCA	3978
Qy	1834	CCTTTGAACCTTTTCGCAAGTGTAGATCCTTATATCCCGAGAACTTTACAGCAACAAT	1893
Db	3979	CCTTTGAACCTTTTCGCAAGTGTAGATCCTTATATCCCGAGAACTTTACAGCAACAAT	4038
Qy	1894	TCAGACTGCTGCCAGAATTTCTCTGGAAGAGCCCTTCTGATTCAGAGAAGGATTTGTCTGG	1953
Db	4039	TCAGACTGCTGCCAGAATTTCTCTGGAAGAGCCCTTCTGATTCAGAGAAGGATTTGTCTGG	4098
Qy	1954	CAAGTCTGATCTAGAGAATTTCTCCAGTCTGGAGGCTTCTCTGA 1998	
Db	4099	CAAGTCTGATCTAGAGAATTTCTCCAGTCTGGAGGCTTCTCTGA 4143	
RESULT 9			
AC107842 215781 bp DNA linear ROD 22-OCT-2003			
LOCUS Mus musculus chromosome 1, clone RP23-30409, complete sequence.			
DEFINITION AC107842			
ACCESSION AC107842.13 GI:37806519			
VERSION HTG.			
KEYWORDS Mus musculus (house mouse)			
SOURCE Mus musculus			
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE 1 (bases 1 to 215781)			
AUTHORS Birren, B., Nusbaum, C. and Lander, E.			
TITLE Mus musculus chromosome 1, clone RP23-30409			
JOURNAL Unpublished			
REFERENCE 2 (bases 1 to 215781)			
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B., Choepel, Y., Collange, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.			
TITLE Direct Submission			
JOURNAL Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
REFERENCE 3 (bases 1 to 215781)			
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,			

Query Match		40.0%;	Score 1175.4;	DB 10;	Length 215781;
Best Local Similarity		72.9%;	Pred. No. 2.7e-306;		
Matches 1642;		Conservative	0;	Mismatches 556;	Indels 53; Gaps 8;
Qy	96	CCATCGCGAGCGCGTGGGGGACTTCCTATCCATGGTGTGAAGCGTCGAGCCGACTAGG	155		
Db	151055	CCGTGCGCGCGCGGAGGGGCTTCTCTATGTTGGAGCGATCTCACCGGCTAGG	151114		
Qy	156	GAACCTCTCTCCC-CCGCAGAGATGAAGTCGATCAGTCGCGCGCTATTTGCGCGGGCTGT	214		
Db	151115	ACGTCCTCTTCCCTAGCCGGGATGACCTAAACCGCGGTGCGACCGCTTGC GCGGGCCTC	151174		
Qy	215	TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTTGCTGCCCTCTGCTGCTTCTTCTGCTGG	274		
Db	151175	TGGGCGCTGCCGTGCAGACTCGTTTGGGAAAGCCGCGCTCTCTGGGCTCTCTTGC	151234		
Qy	275	CTCGAAGATCGCGCCTGGAGCAGCGCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT	334		
Db	151235	CGCGGAATCGACTGTCAGTATCCCACTCCGTGGCTGGCAAGCGGAGACTGTGTAGACC	151294		
Qy	335	TCCTTCGNAATCCGTCGACCTCCAGCGCTGAGCGCGGGGCGCTTACTTGAGAGACTGTC	394		
Db	151295	TCGGATCCAGCTTCGCGCTGACGCGCGCTGAGCTCTGCTCTCTCTCTCTGTTGAGAGCC	151354		
Qy	395	AAGAAAAAGAGATGGAGCGCGGAGCAGCGGATCGCGAAACGGCTTGGCCCTCGGGCG	454		
Db	151355	AAG-GAAGGAGATGGAGCAGGACGACAGGGCCCGGAGCGGCTTGGCCCTCGGCTG	151413		
Qy	455	GGCTTCGGTTCGCGCAACCTTTTTCCTTCGGGATCGCAAGCAGGCTCTTCTAAGTTC	514		
Db	151414	GGCTCTCGTTCGCGCTGCC- --TTCCTTCGGCGATCGCAGCGCTGCTTCTCGAGTTC	151470		
Qy	515	CGGAGCGCTCTTGGCCCGGAAACTCCGGGACCCCACTGCTTTCCTCTGCCAGGCC	574		
Db	151471	CGCGCGCTTCTCTTCGACAAAAATCCCGGAA-----CTCGCTCTGCC	151515		
Qy	575	GAGACTCGGGTCAGTTACTGGAACGAAATGCTCTCCAGCTCTTTCGCGCGCTCCCGGA	634		
Db	151516	GAGCGTCGGACCGACTACTGACCAAAATGCTTCTGAGCTCTTTCGCTCTCCCTAGC	151575		
Qy	635	TTGCTTCAGAAGGTCTAATTTTGGAGCCAACTTTTCGGTGGAAATTTTCGACGATGG	694		
Db	151576	CTAATTCAGAAGCTGCTGCTTTTGGAGCCAGCTTTTTCGGGGGCTGATTTCTTACAGATGG	151635		
Qy	695	CTAGATTTTGTGAGTCTACAGCGCCTTGAGCGCTTGAAGGGACGGGAGAAACGAGCC	754		
Db	151636	CTAGATTTTGGCGAAGTTACAGCGCCTTGAGAGCTTTTGGAGGACGGGAGAACTGCC	151695		
Qy	755	GCCCCACAGCGCAGAAATCTTTGAGTTCGCTGCGAGCTCGACTCTCTCAGACCCCTCGGTC	814		
Db	151696	GCTCCACGCTGCAGAGTCTCTGAGTTCACTGCGGCTGGACTCTTTCGGAAGACTTGGTT	151755		
Qy	815	ACCAGTCCCTTGATTTGGCTAGAGAGGGGATCCACTGGCAATATCTGCCCCCGACCTTA	874		
Db	151756	GTCAGTTCCTCTTGATTTGGCTAGAGAGGGAATCCAGTGGCACTGCTGCTCTCAGACCTG	151815		
Qy	875	AAATTTGAGCTTAAGGCCCAAGGAGTGTCTTTGGACCTCTGACGACAGGCTTTTCTCTTA	934		
Db	151816	GAGTTAAACATCAAGGCCCAAGGAAAGACTTTAGACTCTGACGCGCCACATTTCTCTCTG	151875		
Qy	935	GAGCAGAGCTGTGGGAGTGAGCTGTTGCCAGTAGCTTCAATTCGCGCTGTGTACTCT	994		
Db	151876	GAGCAGAGCTGTGGGAGTGAGGTTGTGTGCCAGTAGCTTCAAGCTGCTTCAAGCTGCT	151935		
Qy	995	AACCGGAACTTGGCTCTTTCGCTTCGGCTCTTAAACATTTCAACGGATAGCAAAATTC	1054		
Db	151936	CACCGAAGACTTGACTCTTCTATCTCTCGGCTCTGAGGGTTTCAAGCTTAGGTAATTC	151995		
Qy	1055	AGTGTGATCCTATTTTGTGAAACCTTCTTACCTGGACTGCTTTCCTAGGCTAGAGTTC	1114		
Db	151996	AAGTAGTTTCTTATCTCTGAAACCTTCTTACCTGGACTACTCTTCCCGAGTTAGGGCTG	152055		

Qy	1115	AGCTATCAGAACACTGATGGAATAGCAGGTAGTCGGCTTCCAGACACTAACCCAGAG	1174
Db	152056	CGCTGTGAGAGCAGCGCTGGAGGTGGCAGTTTGTGGTTTCCGAAACACTAACCCAGAG	152115
Qy	1175	AGCAGCTGCTGAGAGAGGACCATTTGTATCCCGAGCGGCTGAGTGAGAACTCATTTCCG	1234
Db	152116	AGCTGCTATCTTTCTGAAGATGGTGTGACCCCTCAGCGCTTTCGGCGGAGAGATGTCGCA	152175
Qy	1235	GCCTCGTGGCAGGAGTCTCCACCTCTTTCTACGGAAGCGCTTACAGAAAATTCACCATCTT	1294
Db	152176	ACCGCTTGGAGAGGTGTCCGCTCTCTCTACAGAGCGCTTCCGCGAAATCCACACCTT	152235
Qy	1295	CGCATGAAAACGGCTGGAAATTCCTTTCAACAGGCTAACAAAGGGGCAAGATTTACCAACCCCT	1354
Db	152236	CGTATGAAAACGGCTAGAAATTCCT---CCAGGCTAACAAAGGCAAGAGTTTACCCACCCCT	152292
Qy	1355	GACGAGATATGGCTACCAAGCTGAGGAGGAAACAGAGCTTCTCCGATGGATGCCA	1414
Db	152293	GACCAAGATATGGCTATCATAGCTTGGAGAGGAAACATAACCTTCTCCGATGGACCCA	152352
Qy	1415	AAACACTGACAGAGATAACCAACACAGTGTGTTCTCTGCTGGAGACATTTCTCTGAAAC	1474
Db	152353	CAACTTTCACAGATAACCCACAGCGGCTGTCCTCTGCTGACAGAGGCC-----	152405
Qy	1475	ACCAGGAATCCACTGAAGAAAATAAGAAATTAATACTACAGAGGTTCCACTTTGCTTTG	1534
Db	152406	-----GGAGCCCACTGAGAAAACACAGAAATTTGGTGATTCAGAAAGTTTCAAC-----	152452
Qy	1535	GAAGAACAGAGCCCTTCTGAGGCTGTCCATCTAGTCAGATACCTATATGAAAGAGGCGCT	1594
Db	152453	-----AGAGCCCGGAGGAGGAGCTGTGTTTGTGAATTTACCCGTGGAAAAAGAAATGT	152505
Qy	1595	GGAGAGGCGCGAATAAGTGTAGTTGATTTACTCATACCTAGAAGGTGACTTCCCATTTCT	1654
Db	152506	GAAGAGACACACTAATGCACTGACTCTCAGATAGAGAGAGAGGCTTCTCTGTTTCT	152565
Qy	1655	GCCAGACAGCTTGTAGTAACAACTGATAGATATATATTTTGGAGGTGCAATCCAGTGAC	1714
Db	152566	ACCAGACCGTTTGTAGCAACAACTGATAGATATATATTTTGGAGGCGCCCGCCAGTGAC	152625
Qy	1715	CTGGAACAAAGTTCTGATCCAGAGGTGAGGATTTGGATGAGGAGCTGAGGATGATGGT	1774
Db	152626	TTGGAACCGACTCTGATCTGAATCTGAAGTGAAGTTGGGGCGAGGAACTTGAGGACGATGGC	152685
Qy	1775	TTTCATAGTATAGCTCACTGTGACACTCAGACTCAGACCTTGAAACAGACCTTGAAGGCTTTCAC	1834
Db	152686	TTTGAATAGATGCTGCTCTGCTGCTGATCAGACTGAGAACTCGGAAGGCTTTCAC	152745
Qy	1835	CTTTGGAACCTTTCTGAGGTGATGATCTTATATATCCCGAGAACTTTTACAGCAACAAT	1894
Db	152746	CTTTGGAACCTTTTCCACAGTGTAGATCTTACAAACCCCAAACTTTTACAGCCACGAT	152805
Qy	1895	CAGACTCTGCCAGAAATTTCTGGAAGGCTTCTGATTTCCAGAGAGGATTTCTCTGCGC	1954
Db	152806	CAGACGCTGCCAGAAATTTGCCCGCAGAGACCATCAGATTCAGGGAATCTCTGCTGCGC	152865
Qy	1955	AAGTCTGATCTAGAGAAATTCCTCCAGCTCTGGAAGGCTTCTGAGAGCCCTTGAGCATAGT	2014
Db	152866	AGCTGTGCTAGGGA---GCTGTGAGAGGAGCCCTTCCGAGAGACCCCGACCATAGT	152922
Qy	2015	TCTGGGAGGAGATGATCGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGG	2074
Db	152923	TCCGGGAGGAGATGATCGGGAACCGAGTCAGATGAAGCAGAGAACTCTTAAATTTGG	152982
Qy	2075	AACTCATCTGTAATCTGATGACCCCTACACCTTTTAAATTTTAAGGCTCTCTTTCAA	2134
Db	152983	AACTCTTCTGTCAATCTGAGGACCCCTACACCTTTTAAATTTTAAGGCTCTCTTTCAA	153042
Qy	2135	ACATCAGGGGAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTGT	2194
Db	153043	CCGTGAGGAGAAATTGAAAGGCGCTCAGGACTCAAAAGGCTCTCTCTGAGGCTACAGTG	153102
Qy	2195	GCCATTTCTGAGTGTACACCTTACTTTCTTTGTAAGGTGCAGCTGTTGGGAGCCAGAA	2254

```

||||| 153103 GCCTTCTCTGCGCATCATACCTTACTTCTTGTGAAGCCAGCTGTTAGAGAGCAAGAA 153162
||||| 2255 AGTCAATGTCCAGACTCGGTACACGTCGACGTTCTTCTTGGAGGAAGACACACATGTC 2314
||||| 153163 GATTAATGTTCAGAGCTGTGGCTGGTGGTGGCTTCTTGGAGGAAGATACACCATATC 153222
||||| 2315 AAAAGAAAAAGGTAACCTTCTTCTTGAAGAAG 2345
||||| 153223 AAGAGAAAAAGGTTTGTCTTCTTTACATG 153253

RESULT 10
AC128396/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-194B23, *** SEQUENCING IN PROGRESS
***. 6 unordered clones.
ACCESSION AC128396
VERSION AC128396.4 GI:23322373
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 200490)
Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alabrooks,S., Amin,A., Anguiano,D.,
Ayalabechi,V., Ayvagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blewalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gharatne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,R., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaekeleneh,O., Okwunu,G., Olanpunsagoon,A., Pal,S., Parks,K.,
Paaternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajda,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von

```

```

Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Unpublished
2 (bases 1 to 200490)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 200490)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 26, 2002 this sequence version replaced gi:23771040.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYZL
Center clone name: CH230-194B23
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 131649 bases at least Q40
Consensus quality: 136220 bases at least Q30
Consensus quality: 138894 bases at least Q20
Estimated insert size: 161200; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 21048: contig of 21048 bp in length
* 21049 21148: gap of unknown length
* 108746: contig of 87598 bp in length
* 108747 108846: gap of unknown length
* 108847 161819: contig of 52973 bp in length
* 161820 161919: gap of unknown length
* 161920 176640: contig of 14721 bp in length
* 176641 176740: gap of unknown length
* 176741 187347: contig of 10607 bp in length
* 187348 187448: gap of unknown length
* 187448 200490: contig of 13043 bp in length.
*
* Location/Qualifiers
* 1..200490
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-194B23"
* 1..1226
* /notes="wgs contig"
* 13967..15951
* /note="wgs contig"
* 15951..21048
* /note="wgs_contig"
FEATURES
source
misc_feature
misc_feature
misc_feature

```



```

Db 151344 AAGTGAGGATGGGATGAGGAAACCTGAGGATGATGTTGTTGATAGTGATGCGCTCCCTATC 151285
Qy 1798 AGACTCAGACCTTGAACAAGACCTGGAAGGCTTCACTTTGGAACTCTTCTCAGCTGT 1857
Db 151284 TGAATCAGACAGGAGACAGACTCGAAGGCTTCACTTTGGAACTCTTCTCAGCTGT 151225
Qy 1858 AGATCTTATATATCCCGAAGACTTTTACAGCAACAAATTCAGACTGCTCCGAAATTTGTTCC 1917
Db 151224 AGATCTTATCAACCCCGAAGACTTTTACAGCCACAATTCAGACTGCTCCGAAATTTGTTCC 151165
Qy 1918 TGAAGAGCCTTCTGATTCAGAGAGGATTTGTCGGCAAGCTTCATCTAGAGATTCCTC 1977
Db 151164 CAGAGACCAATCTGATTCAGAGAAAGTCTGTCGGCAAGCTTCATCTAGAGAGT---TC 151108
Qy 1978 CCAGTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGA 2037
Db 151107 TCAGGCCACACTCTTCCCGACCCCTGACAGATTCGGGAGGAGAGATGACTGGGA 151048
Qy 2038 ATCTAGTGCAGATGAAGCAGAGACTCTCAAACTGTGGAACCTCATTTCTGTAATTCGTATGA 2097
Db 151047 ATCCAGTGCAGATGAAGCAGAAATCTTAGATTGTGGAACCTCTTCTGTAATTCGTATGA 150988
Qy 2098 CCCCTACACCCCTTAATTTTAAAGGCTCTTTTCAACATCAGCGGAAATGGAAGG 2157
Db 150987 CCCCTACACCCCTTAAATTTTAAAGGCTCTTTTCCAAACGTCAGGAAAGCTGGAAGG 150928
Qy 2158 CTGTCGTGACTCAAGACCCCATCTGATGCTCATTTGTGCCATTTCTGAGTGTACACCTT 2217
Db 150927 CAGTC-----AGGATCTCTGAGGCACAGTGTCTTTCTGGCATCAGACTT 150877
Qy 2218 ACTTTCTTGTAAAGTCAGCTGTTGGGAGCCAAAGATGAATGTCCAGACTCGGTACA 2277
Db 150876 ACTTTCTTGTGAGGCACAACTGTTAGAGAGCAAGAACAAATGTCCAGGCTGTGGCT 150817
Qy 2278 GCGTGACGTTCTTCTGAGGAGAACACACACATGTCAAAGAAAAGGTAACTTCTCT 2337
Db 150816 GGGTGAGGCACTTCTTGAGAAAGATACACCATATCAAGAGAAAAGGTTGTGTTTCCA 150757
Qy 2338 TGAAGAAGTTACTCAGTATTATATATAAGTGGTGATGAGATCGCAAGGA 2386
Db 150756 TTTACATGACTTAAATTTGTTGTTATTTGTTATGTTACGTACGTAGAAGA 150708

```

```

RESULT 12
AC105827
LOCUS
DEFINITION Rattus norvegicus clone CH230-101J7, *** SEQUENCING IN PROGRESS
***
AC105827
VERSION AC105827.3 GI:23265261
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 248458)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisl,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```

```

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,
Loreshwari,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 248458)
Worley,K.C.
Direct Submission
Submitted (10-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 248458)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21736987.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNVL
Center clone name: CH230-101J7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 231489 bases at least Q40
Consensus quality: 232351 bases at least Q30
Consensus quality: 232940 bases at least Q20
Estimated insert size: 249328; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft.html).

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 248458: contig of 248458 bp in length.

FEATURES

source

Location/Qualifiers

1..248458
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-10107"
 6..560
 /note="clone boundary"
 clone end:T7
 site:EcoRI
 end_sequence: BH312302"
 complement(247315..248169)
 /note="clone boundary"
 clone end:Sp6
 site:EcoRI
 end_sequence: BH312303"

misc_feature

misc_feature

ORIGIN

Query Match 38.7%; Score 1137.4; DB 2; Length 248458;
 Best Local Similarity 71.6%; Pred. No. 5.4e-296;
 Matches 1711; Conservative 0; Mismatches 586; Indels 92; Gaps 13;
 1 ATTATGGGCTTCGCTCCACCGCAGCCGCGCTACCGAGTCCCTTCGGGTATCCGGTTG 60
 3992 ATTTTGAGCTTCGCTTCCTTCGCTCTCGCCCGCACCAGCCCTTAACGCTGTGTATTT 4051
 61 CTCAGGGGCTTTTCAACCTCTGTGTCAGTTCGGAACCATCGCCGAGGCGCGTGGGGGACT 120
 4052 CGC-GCCTCGGTGGGCGCGGTCTCCGGAACGTTGCGCGCGCGCGGTGTGAGGGGT 4110
 121 CTTATCCATGTTGAAGCTGTCAGCCGACTAGGGAACCTCTTCCCGCCGCGAGATGGA 180
 4111 CTTTCTATGTGTGAGCGGTCTCACACGACTAGGCGGCTCCATTCCTCGCCGAGATGGA 4170
 181 AGTCGATCAGTCGCGCTATTGCGGGCGTGTCTTCCCTGTCTGCGCGCGCGCTG 240
 4171 CGTAGTTCGGGTTCGCCACCGCTTTCGCGAGCGCCCGCGCTACGCTACGCTGCTA 4230
 241 CGGCATTTCGCTGCTTCGCTTCCTGCTGCTTCGAGTTCGAAATCGGCTTCGAGCAGCGAC 300
 4231 CGGAAGCGCGCTCTCTGGGCC---TCTGCTGGCGCGGATCGGAATCGAATGCAATACCCAC 4287
 301 GCCACGCTGGGCAAGCGCGAGACTCTGTAGCTTCCTCCGAATCCCGTTCGACCTCCAGC 360
 4288 TCCGTGCTGGGCAAGCGCTTAAGACTGTGTAGG---CCTCGATCTCGCCGCTCTGACGC 4344
 361 CGCTGAGCGCGCGCGCTACCTGTCAGAGACTCTCAAGAAAAGGAGATGGAGCCGGGAC 420
 4345 CGCTGAGCTCCGTACTCTGTGCGCGAGAGTTCGCAAGG-AAAGGAGATGGAGCCGGAAC 4403
 421 AGCGGATCGCGAAACGGCTTCGGCCCTCGGCGGGCTTCGGTTCCTGGCCACCCCTTTT 480
 4404 ACACAGGGCTCGGAAGCGGCTTACCTCTCGGATGGGCTCTGTTCCGGCTGCC---TT 4460
 481 CCCTCGCGATCGAAGCAGGCTCTTTAAGTTCCCGAGCGCTCTTGGCCCGGAAACTC 540
 4461 CCTTCGGGCTTCGACGCGCTGCACTTCGAGTTCGGAAGCTTCTCTCGACCAAAATCT 4520
 541 CGGGAACCCCACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTACTGACGAA 600
 4521 CGAG-----GTCTCGCGCGCGCGAGCGCTCGGACCGAGTCTGACGCCAA 4565
 601 ACTGCTCTCCAGCTCTTTGCGCGCGCTCCCGGATTCGCTTCAGAGGTGCTAAATTTGGAG 660

Db 4566 ATTGCTTCCAGCTCCTTGCCCTGCTCCCTAGCTTGTCCAGAGCTGCTTATTTGGAG 4625
 Qy 661 CCAACTTTTCGGTGGATGTTTCCGACAGATGCTAGATTTTCTGCTGAGTCTACAGCGC 720
 Db 4626 TCAGCTTTTCGGTGGCATGATTCCGACAGATGTTAGATTTTTCAGCAAGTTCACAGCGC 4685
 Qy 721 CCTCAGAGCCCTGAAGGGAGCGGGAACACGAGCGCCGCCACAGCGGAGAAATCTTTGAG 780
 Db 4686 TCTAAGAGCTTTGAGAGACCGGAGGAATCAGCGCTCCCACTGAGCAAAATCTTTGAA 4745
 Qy 781 TTCGCTGAGCTCGAC---TCCTCAGACCCCTCGGTCACCACTGCTCCCTTGATTTGGCTAGA 837
 Db 4746 TTCTCTCGGCTGGACTCTTCTTCGGAAGACTTGTGTTGTCTCAGTTCTCTTGTATTGGCTAGA 4805
 Qy 838 GGAGGGGATCCACTGGCAATACTCGCCGCCAGACTTAAATTTGAGGCTTAAAGGCAAGGG 897
 Db 4806 GGAGGGACTCCAGGGGCGAGTGTCTGCTCAGACCTCGGAATTTGAATTTCAAGGCCACGGA 4865
 Qy 898 AAGTGTCTTGGACCTCGCAGCACAGGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGA 957
 Db 4866 AAGAGCTTTAGACTCTACAGCGCACATTTTCTCTTGGAGCAGCAGCTGTGGGAGTGA 4925
 Qy 958 GCTGTTCGCCAGTAGCTTCAATCCCGTCTGTACTCTAAACCGGTCTAGTCTCCACCGAGAACTTGGCTCTTCGCC 1017
 Db 4926 GTTCTGCTCGCGGTAGTCTTCAAAACCGGTCTAGTCTCCACCGAGAACTTGGCTCTTCATC 4985
 Qy 1018 CTCTGGGCTCTAAACATTCACCGCATAGACAAATTTTCAAGGTGTGTATCTTATTTGCTGAA 1077
 Db 4986 CTCTGGGCTCTGAGTACTCAGAGCTTTGGGCAATTTTCAAGGTGTGTATCTTATCTCTGAA 5045
 Qy 1078 CCCTTCTTACCTGAGCTGCTTTCTAGCTAGAGTCTAGCTATCAGAACAGTGTATGAAA 1137
 Db 5046 CCCTTCTTACCTGAGCTGCTTTCCCGAGTTAGAGCTGCGCTGTGCAAGCAGCTGCGAGG 5105
 Qy 1138 TAGCGAGTGTAGTCCGCTTCCAGACACTAAACCCAGAGCAGCTGCTGTGAGAGAGGACCA 1197
 Db 5106 TGGCCAACTTGTGATTTCCGAACTATCCAGAGAGCTGTACCTTTCTGAAGATGC 5165
 Qy 1198 TTGTCTATCCCGCGCTGAGTGCAGAACTCATTTCCGGCTCTGTCGCGAGGATGTCAC 1257
 Db 5166 TTGTCTATCCCGCGCTGAGTGTGCTGAGATGTCCGCAACTGCTGCGGAGGTTGTCGCC 5255
 Qy 1258 TCTTTCTACGGAAGCGCTTACAGAAATTCACCAATTTTCGATGAAAGCGCTGGAATTCCT 1317
 Db 5226 CTTCTCTACAGAGCGCTTCCAGAAATTCACCACTTCCGATGAAACGACTAGATTCCT 5285
 Qy 1318 TCAAAGGCTAACAGGGGCAAGATTTACCAACCCCTGACAGGATTAATGGCTACCAAG 1377
 Db 5286 T---CAGGCAAAACAGGGGCAAGAGTTACCAACCCCTGACCAAGCAATGGCTATCATAG 5342
 Qy 1378 CTTGAGGAGGAAACACAGCGCTTCTCCGATGGAATCCAAACACTGCGAGATACCCCAAC 1437
 Db 5343 CTTGAGGAGGAGCACTGCGCTTCTCCGATGGAATCCGCAACATTCGACAGATAAGGAGC 5402
 Qy 1438 ACAGTTTGTCTCTGCTGAGACATTCCTGGAACACCCAGGAATCCACTGGAAGAAA 1497
 Db 5403 ACAGGAGCTCTCCCTGCTGGAG-----CCAGCCCGCAGCCCTGAGAAAAA 5450
 Qy 1498 AATAAGATTAATACTACAGAGTTCCACTTGTGTTTGGAAAGAGAGCCCTTCTGAGGG 1557
 Db 5451 AGTAGAATGTTGTTGAGGAAATTTTAC-----AGAGCCCGCAGGGAAG 5495
 Qy 1558 CTGTCCATCTAGTCAGATACCTATGGAAGAGGAGCTTGGAGAGGCGCAATAAGTGTAGT 1617
 Db 5496 CAGTCTGTCTGTGAGTTAAGTGTGCAAAAGAAATGTGA----- 5535
 Qy 1618 TGATTAATCTATACCTAGAGGTGACTTCTCCCATTTCTGCCAGACAGCTTGTAGTAAACA 1677
 Db 5536 -GACTACTCAGACATAGGGGTAACTTCCAGTCTCTACAGACAGCTTGTACCAACA 5594
 Qy 1678 ACTGATAGATTAATTTTGGAGGTGATCATCAGTGTGAGCCTGGAACCAAGTTCGTATCAGA 1737

```
Db 5595 ACTGATAGATTATATTTTGGGAGGAGCCTCCAGTCACTTGGACACCAGCTCCGATTCCTGA 5654
Qy 1738 AGGTGAGATTGGGATGAGGAAGCTGAGCATGATGTTTTCATAGTATGATCACTGTC 1797
Db 5655 AAGTGAGATTGGGATGAGGAAGCTGAGCATGATGTTTTCATAGTATGATCACTGTC 5714
Qy 1798 AGACTCAGACTTGAACAAGACCCCTGGAAGGCTTCACTTTTGGAACTCTTTCTCGAGTGT 1857
Db 5715 TGAATCAGACAGGGAACAAGACTCCGAAGGCTTCACTTTTGGAACTCTTTCTCGAGTGT 5774
Qy 1858 AGATCCTTATATCCCAAGACTTTACAGCAACATTCAGACTCTGCGCAATTTGCTTC 1917
Db 5775 AGATCCTTATACACCCCAAGACTTTACAGCAACATTCAGACTCTGCGCAATTTGCTTC 5834
Qy 1918 TGAAGAGCCTTCTGATTCAGAGAAGGATTTCTGCGCAAGTCTGATCTAGAGAATTCCTC 1977
Db 5835 CAGAGACCCATCTGATTCAGAGAAGTCTGCTGCTGCGCAACTCTGATGTAGGAGT---TC 5891
Qy 1978 CCAGTCTGGAAGCCTTCTGAGACCCCTGAGCATGATTTCTGGGAGGAAGATGCTGGGA 2037
Db 5892 TCAGGCCACACTCTCTCCCAAGCCCTGACCAAGTTCGCGGAGGAGGATGACTGGGA 5951
Qy 2038 ATCTAGTGCAGATGAGCAGAGACTCTCAAACTGTGGAACCTCATCTGTAATTCGTATGA 2097
Db 5952 ATCCAGTGCAGATGAGCAGAGAATCTTAGATTTGGAACCTCTTTCTGTAATTCGTAGGA 6011
Qy 2098 CCCCTACAACTTTAAATTTTAAGGCTCTCTTTTCAAACTCAGCGGGAATGAGAAGG 2157
Db 6012 CCCCTACAACTTTAAATTTTAAGGCTCTCTTTTCAAACTCAGCGGGAAGCTGGAAGG 6071
Qy 2158 CTGTCTGAGTCAAGACCCCATCTGAGTGCATTTGTGSCCATTTCTGAGTGTCAACCTT 2217
Db 6072 CAGTC-----AGGCATCTCTGAGGCCACAGTGGTCTTTCTGGGCATCAGACCTT 6122
Qy 2218 ACTTCTTGTAAAGTGCAGCTTTGGGAGGACCAAGATGTAATGTCAGACTCGGTACA 2277
Db 6123 ACTTCTTGTGAGGCACAACTGTTAGAGAGCAAGAGCAATTTGTCCAGGCTGTGGCT 6182
Qy 2278 GCGTGACCTTCTTCTGGAGGAAGACACACATGTCAAAAGAAAAGGTAACCTTCTCT 2337
Db 6183 GGGTGAGGCATTTCTGGAGGAAGATACACCATATACAGAGAAAAGGTTTGTTCCTCA 6242
Qy 2338 TGAAGAAGTTACTGAGTATATATAAGTGGTGTGATGAGGATCGCAAGA 2386
Db 6243 TTTACATGACTTATAATTTGTTGTTATTTGATATGTTTACAGTAGAGA 6291

RESULT 13
AC119774
LOCUS Rattus norvegicus clone CH230-146P11, DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-146P11, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC119774
VERSION AC119774.6 GI:25095262
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 212261)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Avodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Bigswain, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
```

```
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Loulseg, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojars, A., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaja, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 212261)
Worley, K. C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 212261)
Rat Genome Sequencing Consortium.
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23194778.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVQB
Center clone name: CH230-146P11
----- Summary Statistics
```

```

Assembly program: Phrap; version 0.990329
Consensus quality: 198836 bases at least Q40
Consensus quality: 201017 bases at least Q30
Consensus quality: 202375 bases at least Q20
Estimated insert size: 203079; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11068: contig of 11068 bp in length
* 11069 11168: gap of unknown length
* 11169 212261: contig of 201093 bp in length.
FEATURES             Location/Qualifiers
     source            1..212261
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-146P11"
     misc_feature      1..1217
                     /note="wgs_contig"
ORIGIN
Query Match      38.1%; Score 1120.2; DB 2; Length 212261;
Best Local Similarity 71.6%; Pred. No. 2.4e-291;
Matches 1771; Conservative 0; Mismatches 603; Indels 99; Gaps 19;

QY 347  CGTGACCTCCAGCGCTGAGCGCGCGCCCTACTCTGAGACTGTGCAAGAAAAGAG 406
Db 140598  CTTGCTCTGATGCGCGCTGAGCTCGGTACTCTGCGCGGTGAAGTTGCCAAG- 140656

QY 407  ATGGAGCGGGACAGCGGATCGCGAAAGCGCTTGGCCCTCGCGCGGCTTCGGTTC 466
Db 140657  ATGGAGACGGGAACACAGGGCTC-AMAGCGCCTACTCTCGGATGGCTCTCGATTA 140715

QY 467  TGGCCACCCCTTTTCCCTCGGCGATCGCAAGAGGCTCTTCTAAGTTCCCGACGCTCTT 526
Db 140716  TGCTGCGCTTCTTAG---GGCGATCACAGCGCTGACCTTCTGAGTTCCCGAAACCTTCT 140772

QY 527  GGCCCGGAAACTCGCGGAACCCACATGCTTTCTGCGCCACCGCGAGACTCGGTC 586
Db 140773  TCTCAACCAAATCTCGAGG-----CTGTGCTGTGTCGAGCCTCAGACC 140817

QY 587  AGTTACTGGACGAACCTGCTCTCCAGCTCCTTTGCGCGCTCCCGGATTTGCTTCAGAG 646
Db 140818  AGGTACTGGACCAATTTGCAATTCAGATTCCTGTG---CTCTCCCTAGCTTGTCTCAGAG 140874

QY 647  GTGCTAAATTTGAGGCAACTTTTTCGGTGAATGTTTCCGACAGATGGCTAGATTTGCT 706
Db 140875  CTGCTTAATTTGAGTCAGCTTTTCGGTGCATGATCTGACCAAGTGGCTATACCTTTCA 140934

QY 707  GGAGTCTACAGCGCCCTGAGACCTTGAAGGACGGGA-GAAACAGCGCGCCCGACAGC 765
Db 140935  GCAAGTTACAGCTCCCTCGAGCTTTGAGAGACCGGAGGAATCAGCGCGCTCCACGTT 140994

QY 766  GCAGAAATCTTTGAGTTGCTGCAGCTCGA---CTCCTCAGACCCCTCGGTACCAGTCC 822
Db 140995  GCAGAAATCTTTGAGTTCTCTCGGCTGGAATTTTCTCGGAAGACTTTGGCTGTCAAGTTC 141054

QY 823  CTTGATTTGGCTAGAGGGGATCCACTGGCAATTAATCTGCGCCCGACAGACCTTAAAT 882
Db 141055  TCTGATCGCTAGAGGACGGACTCCAGTGGCAGTGCTCGTCTCAGACCTTGAATTTGA 141113

QY 883  GCTTAAGGCCAGGGAAGTGTCTTGACCCCTGACAGCAGGCTTTTCTTTAGAGCAGCA 942
Db 141114  ACTCAAGGCCAGGAAAGAGCTTTTAGACTCTACAGCGCACACTTTCTCCTCATGGAGCAGCA 141173

```

```

QY 943  GCTGTGGGAGTGGAGCTGTTGCCAGTAGCCTTCAATCCCGTCTGTACTCTTAACCCGGA 1002
Db 141174  GCTGTGGGAGTGGAGTGTGTCGCCCGCTCTCTTCAACCGATAGTCTCTCCACCAAGA 141233

QY 1003  ACTTGGCTCTTTCGGCCTCTGGGCTCTTAAACATTTCAAACGATAGACAATTTCACTGGT 1062
Db 141234  ACTTGGCTCTTCACTCTCTGGGCTCTGAGCACTCAGAGCTTGGGCAATTTCAAGGTAGT 141293

QY 1063  ATCCTATTGTGTAACCTTCTTACCTTGGAGTGTCTTCTTAGCTAGAGTCACTATCA 1122
Db 141294  TTTCTATTCTGTAACCTTCTTACCTTGGAGTGTCTTCTTCCAGTTAGAGTGTGCTGCA 141353

QY 1123  GAACAGTGTGGAATAGCGAGGTAGTGGCTTCCAGACACTTAACCCAGAGAGCAGCTG 1182
Db 141354  GAACAGCATCTGAGGTGGCCACTTGTGGATTTCCGAGCGCTATCCCGAGAGACTGTTA 141413

QY 1183  CCTGAGAGAGACCATTTGTATCCCGACCGCTGAGTGCAGAACTCATTCGGGCTCGTG 1242
Db 141414  CTTTCTGAAAGTGTGTTGTCA-CTCCAGCGCTGCGTGTGAGATGTGCGCAACTGCTG 141472

QY 1243  GCAGGATGTCCACTCTTTCTAGGGAAGGCTTACCGAATTTCAACATCTTCGCATGAA 1302
Db 141473  GCGAGGTGTCCAGCCCTCTCTACAGAAAGGCTGCGCGAATTCACCACTTCGGATGAA 141532

QY 1303  ACGGCTGGAATTCCTTCAACAGGCTAAACAGGGGCAAGATTTACCCACCCCTGACCAAGA 1362
Db 141533  ACGGCTAGATTTCTTTC-----AGAGCAAGAGTTACCAACCCCTGACCAAGA 141580

QY 1363  TAATGGCTACCAAGCTGAGGAGGAACACAGCCTTCTCGGATGGATCAAAACACTG 1422
Db 141581  CAATGGCTATCATAGCTTGGAGGAGGAGCACTGCTTCTCGGATGGACCGCAACATTTG 141640

QY 1423  CAGAGATAACCCACACAGATTTGTTCTGCTGCTGGAGACATTTCTCTGGAACACCCAGGA 1482
Db 141641  CACAGATAAGGTAGCACAGGCTATGTCCCTGCTGGAG-----CCAGCCCGCA 141688

QY 1483  ATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTTCCACTTCTTTTGGAGAAGA 1542
Db 141689  GCCCACTGAGAAAAAAGTAGAATTTGGTGTGAGGAAGTTTCAC-----A 141733

QY 1543  GAGCCCTCTTGGAGGCTGTCCATCTAGTGAGATACCTATGGAAGAGAGCTCGAGAGGG 1602
Db 141734  GAGCCCGCAGGAAGCAGCTGTCTGTGAGTTAAGTGTGCAAAAAGAAATGTGAAGAGGA 141793

QY 1603  CCGAATAGTGTAGTTGATTAATCATACCTAGAGGTGACCTCCCATTTCTGCGACAGC 1662
Db 141794  CCACACTAATGTAATGACTACTCAGACATAGGGGGAAACCTTCTCTGTTTCTCCGACCC 141853

QY 1663  AGCTTGTAGTAAACAAACTGATAGATTATATTTTGGGAGGTGTCATCCAGTGCACCTTGA 1722
Db 141854  AGCTTGTACCAACAACTGATAGATTATATTTTGGGAGAGGCTCCAGTGCATTTGGACAC 141913

QY 1723  AAGTTCTGATCCAGAGGCTGAGGATGGGATGAGGAAGCTGAGGATGATGTTTGTATAG 1782
Db 141914  CAGCTCCGTTTCTGAAAGTGGAGATTGGGATGAGGAACCTGAGTATGATGTTTGTATAG 141973

QY 1783  TGATAGCTCACTGTGACACTCAGACCTTGAACAGACCCCTGAAGGGCTTCACTTTTGGAA 1842
Db 141974  TGATGGCTCCCTGTCTGAAATCAGACAGGGAACAAGATCCGAGGGCTTCACTTTTGGAA 142033

QY 1843  CTCCTTCTGCAAGTGTAGATCCTTATAATCCCGAAGACTTTTACAGCAACAATTCAGACTGC 1902
Db 142034  CTCCTTCTACAGTGTAGATCCTTATAACCCCGCAACACTTTACAGCCACGATTCAGACTGC 142093

QY 1903  TGCAGAAATTTGTTCTGAGAGGCTTCTGATTGAGGAAGGATTTGTCTGCGAAGTCTGA 1962
Db 142094  TGCAGAAATTTGCCCGCCAGAGACCCATCTGATTCAGGAAGTCTCTGGTCTGTGGCAACCCGGA 142153

QY 1963  TCTAGAGAAATTCCTCCAGCTGTGGAAGCCTTCTCTGAGACCCCTGAGCATATGTTCTGGGA 2022
Db 142154  TGTAGGGA---GGTCTCAGGGCACACTCTCTTCCGAGACCCCTGACAGAGTTTCTGGGA 142210

```

```
QY 2023 GGAAGATGATCGGGAACTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAACTCAT 2082
Db 142211 GGAGGATGACTGGGAACCTAGTGCAGAGGAAGCAGAAAATCATAGATGTGGAACTCTCT 142270
QY 2083 CTGTAAATCTGATGACCCCTACACCCCTTTAAATTTTAAAGGCTCTTTTCAAAATCAGG 2142
Db 142271 CTGTAAATCTGAGGACCCCTACACCCCTTTAAATTTTAAAGGCTCTTTTCCAAAGCTTAGG 142330
QY 2143 GGAATAATGAGAAAGGCTGCTGCTGACTCAAGACCCCATCTGAGTCCATTTGTGGCCATTTTC 2202
Db 142331 GAAGAGCTGGAAGAGGCTGTGAGGACTCGAAGGCATCTCTGCGGCGCAGTGTCTTCTGTC 142390
QY 2203 TGAGTGTACACCTTACTTTCTTTGTAAGTGC-----AGCTGTGGGGAGCAAGAAAGT 2257
Db 142391 TGGGCATCAGACCTTACTTTCTTTGTAAGGCAAGAAATTTAAATAGAGAGCCCAAGAGA 142450
QY 2258 GAATGTCAGACTCGGTACAGCGTGACGCTTCTTTCTGAGAGCAAGACACACATGTCAAA 2317
Db 142451 TAAATGTAGGCTGTGGGCTGGGTGAGGCTCTTTTGTGGAAGATACACCCATATCAAG 142510
QY 2318 AGAAAAAAGGTAACTCTCTTGAAGAGCTTACTGAGTATTATATAAGTGGTGATGAGGAT 2377
Db 142511 AGAAAAAAGATAACTCTCTTGAAGAGCTTACTGAGTATTATATAAGTGGTGATGAGGAT 142570
QY 2378 CGCAAGGACCATGGGAAGATTTGCAAGGATGATGATGATGATGATGATGATGATGATGAT 2437
Db 142571 CGCAAGGATTAATGGGAAGATTTGCAAGGATGATGATGATGATGATGATGATGATGAT 142630
QY 2438 GAAACAGAGATGCTATTTGGATATTTGCTTGTGATGATGATGATGATGATGATGATGAT 2497
Db 142631 GAAACAGAGATGCTATTTGGATATTTGCTTGTGATGATGATGATGATGATGATGATGAT 142686
QY 2498 AGACTCCAGGGAACATGCTTCAAGAGGACTTAATGTTCTCAAGCAATGTTGAGTTG-GCAG 2556
Db 142687 AGACTGCA-AGACCGTGTTCAGAGGACTTCTGTTGTACAGAGTGTCTAGCAGAGGGA 142745
QY 2557 CCTGAGTCTAGCTAGCATACATCTTCTGATGAGAGTGTCTTTTAAACCAAAAT 2616
Db 142746 CGAGCTGCGAGCCCTCACCCACTCCATCTCTTACTTGAGAGGTTCTCTTATAGCAAAAC 142805
QY 2617 CTGGCAGCTGCTCTTGCATTTTCTTTTGTAGGAAATGTAACCTGATCTAGTTTA 2676
Db 142806 ACTGCGAGCTGCTCTTGCATTTTCTTTTGTAGGAAATGTAACCTGATCTAGTTTA 142865
QY 2677 ATTTTCTTTTGTCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATA 2736
Db 142866 AGTTATTTTGTGATGATGTCTCA-----TTAGAAACACACACATCTTAATA 142914
QY 2737 ATGAAGGATGAACCTAGTGTGATTTCTAATCTCCCTTTTGTAGTTAGTGTGCTT 2796
Db 142915 ATGAAGGAT-----CTCTAATCTTAACTCTTTCTTTTCTATATAGCTGGATGGGT 142968
QY 2797 TTAATGTCTTTT 2809
Db 142969 TTTTCTTTTCTTTT 142981

RESULT 14
AC094236/c
LOCUS AC094236 240836 bp DNA linear HTG 14-NOV-2002
DEFINITION Rattus norvegicus clone CH230-3B24, *** SEQUENCING IN PROGRESS ***
4 unordered pieces.
ACCESSION AC094236
VERSION AC094236.5 GI:24958472
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 240836)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alibrooks,S., Amin,A., Anguiano,D.,
```

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewick, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabris, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gbrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulesged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekemele, O., Okwunonu, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L., L., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 240836)
Worley, K. C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 240836)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (14-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 14, 2002 this sequence version replaced gi:22772949.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAGF
Center clone name: CH230-3B24
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 194130 bases at least Q40
Consensus quality: 198891 bases at least Q30
Consensus quality: 202078 bases at least Q20
Estimated insert size: 194853; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 184581: contig of 184581 bp in length
* 184582 184681: gap of unknown length
* 184682 238285: contig of 53604 bp in length
* 238286 238385: gap of unknown length
* 238386 239477: contig of 1092 bp in length
* 239478 239577: gap of unknown length
* 239578 240836: contig of 1259 bp in length.

FEATURES

source

1..240836
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-3B24"
1984..3036
misc_feature
/note="wgs_contig"

ORIGIN

Query Match 33.7%; Score 991.4; DB 2; Length 240836;
Best Local Similarity 72.3%; Pred. No. 1.9e-256;
Matches 1494; Conservative 0; Mismatches 496; Indels 75; Gaps 13;

QY 754 CGCCCCCAGCGCAGAAATCTTTGAGTTTCGCTGCGAGTCGGA---CTCCTCAGACCCCTC 810
DB 174258 CGCTCCACGTTGCAGAAATCTTTGAGTTTCGCTGCGAGTCGGAATTTTCTTCGGAAGACTT 174199

QY 811 GGTCAACAGTCCCTTGTGCTAGAGAGGGGATCCACTGGCAATACTGCGCCCGCCAG 870
DB 174198 GGCTGTCACTTCTTGTGCTAGAGAGCGGACTCCAGTGGCAGTGTCTGCTCAG 174139

QY 871 CTTAAATTTGAGACTTAAGGCCAAGGAGTCTTTTGGACCTCGACACAGGCTTTTCT 930
DB 174138 CTTGGAAT-TGAACCTCAAGGCCCGAAGAGACTTTAGACTCTACAGGCGACACTTTCT 174080

QY 931 CTTAGACGACGAGCTGTGGGAGTGGAGCTGTGCGCAGTAGCTTCAATCCCGTCTGTA 990
DB 174079 CATGAGCAGCAGCTGTGGGAGTGGAGTTGTGCTGCCCCCTCTTTCAACCGATCTAGT 174020

QY 991 CTCCTAACCGGAACTTGGCTCTTGGCCCTCTGGGCTCTTAAACATCTCAACGATAGACA 1050
DB 174019 CTCCTAACCGGAACTTGGCTCTTGGCCCTCTGGGCTCTTAAACATCTCAACGATAGACA 173960

QY 1051 TTTTCAGTGTGATCTCTATTTCTGTAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1110
DB 173959 TTTTCAGTGTGATCTCTATTTCTGTAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 173900

QY 1111 AGTCAGCTATCAGAACAGGTGATGGAATAGCGAGGTAGTCGGCTTCCAGACACTAACCC 1170

Db	173899	GCTCGCTGTTCAGAAACAGCATCTGAGGTGGCCAACTTTGTGATTTCCGAGCGCTATCCCC	173840
QY	1171	AGAGAGCAGCTGCTGAGAGAGGACATTTGTCTATCCCGACCGCTGAGTGCAGAACTCAT	1230
Db	173839	AGAGAGCTGTTCATCTTTTGAAGGTGTGTGTCA-CTCAGCGCTGCTGCTGAGATGTC	173781
QY	1231	TCGGGCTCGTGGCAGGGATGTCACCTCTTTCTACGGAAGGCTTACAGAAATTCACCA	1290
Db	173780	GGCAACTGCTGGCGGAGGTGTCCAGCCCTCTCTACAGAGGCTTCCGGAATTCACCA	173721
QY	1291	TCCTTCGATGAAACCGCTGGAATTTCTTCAACAGGCTTAACAAGGGGCAAGATTTACCCAC	1350
Db	173720	CATTTCGATGAAACCGCTAGAAATTCCTTC-----AGAGGCAAGAGTTTACCAAC	173673
QY	1351	CCCTGACAGGATTAATGGCTTACACAGCCTGGAGGAGAACACAGCTTCTCCGGATGGA	1410
Db	173672	CCCTGACAGGATTAATGGCTTATCATGCTTGGAGGAGGACACTGCTTCTCCGGATGGA	173613
QY	1411	TCCAAACACTGACAGATTAACCCCAACACAGTTTGTTCCTGCTCTGAGACATTCCTGG	1470
Db	173612	CCCGCAACTTGCACAGATTAAGGTAGCACAGGCTATGTCCCTGCTGGAG-----	173563
QY	1471	AAACACCCAGGAATCCACTGAAAGAAATAAGAAATTAATACTACAGAGTTCCACTTGC	1530
Db	173562	--CCAGCCCGCAGCCCACTGAGAAAAAAGTAGAATTCGTGTTGAGGAAGTTTCAC----	173509
QY	1531	TTTGGAGAGAGAGCCCTTCTGAGGCTGTCCATCTAGTCAGATACCTATGGAAGGA	1590
Db	173508	-----AGAGCCCCCAGGGAAGCAGTCTCTCTGTGAGTTAAGTTGCAAAAGA	173460
QY	1591	GCCTGGAGAGGGCCGAATTAAGTGTAGTTGATTACTCATCTAGAAAGTGCCTTCCCAT	1650
Db	173459	ATGTGAAGAGGACCACTAATGTAAATTGACTACTCAGACATAGGGGGAACCTTCTGT	173400
QY	1651	TTCTGCCAGACCACTTGTGTAGTAACAACTGATAGATTAATTTTGGAGGTGCATCCAG	1710
Db	173399	TTCTCCAGACCACTTGTGTACCAACAACTGATAGATTAATTTTGGAGGAGCTCCAG	173340
QY	1711	TGACCTGGAAACAGTTCTGATCCAGAGGTTGGAGTGGAGTGAAGCTGAGGATGA	1770
Db	173339	TGACTTGGACACCACTCCGCTTCTGAAAGTGGAGTTGGAGTGGAGAACCTGAGTATGA	173280
QY	1771	TGTTTTGTATGATGATAGCTCACTGTGAGCTCAGACCTTGAAACAAGCCCTGAAGGCT	1830
Db	173279	TGTTTTGTATGATGATGCTCCCTGTCTGAATCAGACAGGGAACAAGACTCCGAAGGCT	173220
QY	1831	TCACCTTTGGAACTCTTTTCTGAGTGTAGATCTTTAATAATCCCGCAGAACTTTACAGAAC	1890
Db	173219	TCACCTTTGGAACTCTTTTCTACAGTGTAGATCTTTAACAACCCCAACACTTTACAGCCAC	173160
QY	1891	AATTCAGACTGCTGCCAGAAATTTCTTGAAGAGCTTCTGATTCAGAGAGGATTTGTCT	1950
Db	173159	GATTCAGACTGCTGCCAGAAATTTGCCCAAGAGACCCATCTGATTCAGAGAGTCTGTGTC	173100
QY	1951	TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAGGCTTCTGAGACCCCTGAGCA	2010
Db	173099	TGGCAACCCCGATGTAGGGA---GGTCTCAGGCCACACTCTCTCCGAGAGCCCTGACCA	173043
QY	2011	TAGTTCTGGGAGGAAAGATGATCTGGGAATCTAGTGGAGATGAAGCAGAGACTCTCAAACT	2070
Db	173042	GAGTTCTGGGAGGAGGATGATCTGGGAACCTTAGTGGCAGAGGAAGCAGAAAAATCATAGAT	172983
QY	2071	GTGGAATCTATCTGTAAATCTGTATGATGATCCCTCAACCTTTTAAATTTTAAAGGCTCTTT	2130
Db	172982	GTGGAATCTCTCTGTAAATCTGTAGGACCCCTCAACCTTTTAAATTTTAAAGGCTCTTT	172923
QY	2131	TCAAACTCAGGGGAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTAGTCCAT	2190
Db	172922	CCAAACGTTAGGGAAGAGCTGGAAGGCTGTGAGGACTCGAAGGCTCTCTGGGGCCAC	172863
QY	2191	TGTGGCAATTTCTGAGTGTACACCTTACTTCTTGTGTAAGGTGC-----AGCTGTGGGG	2245

```
Db 172862 AGTGGTCTTGCTGGGCATCAGACTTACTTCTTTGTAAGCCACAAGAACTTAATTAGA 172803
Qy 2246 AGCCAAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGCTTCTTCTTGGAGGAAGACAC 2305
Db 172802 GAGCCAAAGAGATAATTTGAGGGCTGTGGGCTGGGTGAGCTCTTTTGGTGAAGATAC 172743
Qy 2306 ACACATGTCAAAAGAAAAGGTAACTTCTTCTTGAAGAGTACTGAGTATATATAAGT 2365
Db 172742 ACCCATATCAAGAGAAAAGATAAACCCTTCTTGAAGAGTACTGAGTATATATAAGT 172683
Qy 2366 GGTGATCGAGATCCCAAGGACCAATGGAGAGATTTCCAGGGAGTGAATCAGGTCCAG 2425
Db 172682 GGTGATGAGGCTCCCAAGGAGTATGGAGAGAAATTTGCAAGGGATGGATGCAAGTCCAG 172623
Qy 2426 AAACGAATTCAGAGAAACAGAGATGCTATTGGATATTGCTTGACATTTGACATTTGAACACAGAGAA 2485
Db 172622 AAACGAATTCAGAGAAACAGAGATGCTATTGGTACTGCTTGAGCTTTGAGCAC----AA 172567
Qy 2486 AGAATGTTTAATAGACTCCAGGGAACATGCTTTCAAAGGACTTAATGTTCTCAAGCAATGT 2545
Db 172566 AGAGTGTTTAGTAGACTGC-GAGACCGTGTGTTCAAAGGACTTCTGTTGTACAGCAGTGC 172508
Qy 2546 TGAGTTG-GCAGCCTGTAGTCTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTT 2604
Db 172507 TAAACAGAGGAGCAGCTGCAGCCCTCACCCACTCCATCTCTTACTTGTGAGAGGGTTCTC 172448
Qy 2605 TTAAGAAACAAATCTTGGCAGCTGCTTGTGACATTTTCTTTTATGAGGAAATGTAACCTT 2664
Db 172447 TTATAGACAAACACTGCGAGCTGCTTGTGNCATGGGTTTAAAGAAACGACTTGGTTCTA 172388
Qy 2665 GGATCTAGTTAAATTTTTTTTTTTTGGCAACATATCCCACTCAGAAACATTCAGGTTTGAAG 2724
Db 172387 GTGATGTAGTTAGTTATTTTGGATAATGTGTCTCA-----TTAGAAACACAC 172339
Qy 2725 CCAGCCCTGATAATGAGGATGAAGTACTAGTGTGATTTCTTAATCTCCCTTTTGTGATTAG 2784
Db 172338 ACAACTCTAAATGAGGGAT-----CTCTAATCTTAAATCTCTTCTTCTTCCATATATAG 172285
Qy 2785 TTGGATGTGCTTTTAAATGTCTTT 2809
Db 172284 CTGGATGTGGTTTTTTTTTTTTTTT 172260

RESULT 15
BD148615 772 bp DNA linear PAT 17-JAN-2003
LOCUS BD148615
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD148615
VERSION BD148615.1 GI:27854373
KEYWORDS JP 2002191363-A/3458.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota,T., Teogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Iehli,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Oteuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 3458 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/3458
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOHAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
```

```
FEATURES
source
Location/Qualifiers
FT source 1..772
/organism="Homo sapiens (human)".
Location/Qualifiers
1..772
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 25.1%; Score 738; DB 6; Length 772;
Best Local Similarity 98.8%; Pred. No. 4.2e-188; Indels 1; Gaps 1;
Matches 752; Conservative 0; Mismatches 8;
Qy 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTACCCAGTCTTCCGATATCGGTTG 60
Db 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTACCCAGTCTTCCGATATCGGTTG 60
Qy 61 CTCAGGGGCTTTTCAACCCCTCTGTCAAGTGGAAAAACATCGCCGAGGCGGTGGGGGACT 120
Db 61 CTCAGGGGCTTTTCAACCCCTCTGTCAAGTGGAAAAACATCGCCGAGGCGGTGGGGGACT 120
Qy 121 CCTATCCATGCTGTTCAAGCGTCCGACCGACTAGGGAACCTCTTCCCGCCAGGATGGA 180
Db 121 CCTATCCATGCTGTTCAAGCGTCCGACCGACTAGGGAACCTCTTCCCGCCAGGATGGA 180
Qy 181 AGTCGATCAGTCGCGCCCTATTGCGGGGCTGTTCTTCCCTGTGTTCTGCCGCCGCTG 240
Db 181 AGTCGATCAGTCGCGCCCTATTGCGGGGCTGTTCTTCCCTGTGTTCTGCCGCCGCTG 240
Qy 241 CCGCATTCGCTGCGCTCTGTGGCTTTTCTGTGGCTGGAAGATCGGCTCGAGAGCGGAC 300
Db 241 CCGCATTCGCTGCGCTCTGTGGCTTTTCTGTGGCTGGAAGATCGGCTCGAGAGCGGAC 300
Qy 301 GCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360
Db 301 GCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360
Qy 361 CGCTGAGCGCGCGGCTTCTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGGAC 420
Db 361 CGCTGAGCGCGCGGCTTCTACCTGAGAGACTGTCAAGAAAAAGAGATGGAGCGGGGAC 420
Qy 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGGCTTCCGGTCTTGGCCACCTTTT 480
Db 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGGCTTCCGGTCTTGGCCACCTTTT 480
Qy 481 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCGAGCGCTTCTGGCCCGGAAACTC 540
Db 481 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCGAGCGCTTCTGGCCCGGAAACTC 540
Qy 541 CGGGAACCCCACTGCTTCTCTGCCAGCCCGAGACTCGGGTCAAGTTACTGGAGCAA 600
Db 541 CGGGAACCCCACTGCTTCTCTGCCAGCCCGAGACTCGGGTCAAGTTACTGGAGCAA 600
Qy 601 ACTGCTCTCCAGCTCTTGGCGGCTTCCCGGATTTGCTTCAGAAAGTGTCTAATTTGGAG 660
Db 601 ACTGCTCTCCAGCTCTTGGCGGCTTCCCGGATTTGCTTCANAAAGTGTCTAATTTGGAG 660
Qy 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCGAGATGCTAGATTTTGTGGAGTCTAC-AGCG 719
Db 661 CCAACTTTTTCGGTGGAAATGTTTCCGACCGAGATGCTAGATTTTGTGGAGTCTACAGCG 720
Qy 720 CCCTGAGAGCCCTCAAGGAGCGGAGAAACCCAGCCGCCCC 760
Db 721 CCCTGAGAGCCCTCAAGGAGCGGAGAAACCAAGCCGCCCC 761

Search completed: September 16, 2005, 01:57:34
Job time : 12271.5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:34:33 ; Search time 22770.5 Seconds
(without alignments)
11635.782 Million cell updates/sec

Title: US-10-650-482-3
Perfect score: 5468
Sequence: 1 cggctctcgtctgcctg.....attaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ets.*
12: gb_ey.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5234.2	95.7	5367	10	BC058078 Mus muscu
2	3115	57.0	3704	10	BC006897 Mus muscu
3	3100	56.7	215781	10	AC107842 Mus muscu
4	2412.4	44.1	212261	2	AC119774 Rattus no
5	2188.8	40.0	240836	2	AC094236 Rattus no
6	1689.2	30.9	200490	2	AC128396 Rattus no
7	1689.2	30.9	227054	2	AC098956 Rattus no
8	1689.2	30.9	248458	2	AC105827 Rattus no
9	1320.8	24.2	2942	6	BD160691 Primer fo
10	1320.8	24.2	2942	6	AX884059 Sequence
11	1320.8	24.2	2942	9	AK027650 Homo sapi
12	1320.8	24.2	5264	9	BC065280 Homo sapi
13	1124.8	20.6	167079	9	AL606489 Human DNA
14	1124.8	20.6	328863	2	AL627202 Homo sapi
15	671	12.3	166233	2	AC048369 Homo sapi
16	489.4	9.0	257938	2	AC124919 Rattus no
17	353	6.5	263621	2	AC095784 Rattus no
18	262.6	4.8	212115	2	AC116701 Mus muscu
19	239.2	4.4	772	6	BD148615 Primer fo

AX868553	Sequence	6	AX868553	4.4	772	6	AX868553
AC048369	Homo sapi	2	AC048369	3.8	166233	2	AC048369
G53188	SHGC-84724	11	G53188	3.7	348	11	G53188
AC112836	Rattus no	2	AC112836	3.4	303230	2	AC112836
AX401620	Sequence	6	AX401620	3.3	540	6	AX401620
AC127784	Rattus no	2	AC127784	2.1	268805	2	AC127784
AC127924	Rattus no	2	AC127924	2.1	298235	2	AC127924
BD020531	Novel gen	6	BD020531	2.1	689	6	BD020531
BD100469	Novel gen	6	BD100469	2.1	689	6	BD100469
BD153658	Primer fo	6	BD153658	2.0	543	6	BD153658
AX873596	Sequence	6	AX873596	2.0	543	6	AX873596
Continuation (2 of				2.0	110000	2	AC120739_1
AC098930	Rattus no	2	AC098930	1.9	252921	2	AC098930
AC142019	Rattus no	2	AC142019	1.9	89435	2	AC142019
AC133740	Rattus no	2	AC133740	1.9	211582	2	AC133740
AC105660	Rattus no	2	AC105660	1.9	240626	2	AC105660
AC137459	Rattus no	2	AC137459	1.9	189173	2	AC137459
AC123126	Rattus no	2	AC123126	1.9	231511	2	AC123126
AC139951	Rattus no	2	AC139951	1.9	194401	2	AC139951
AC142069	Rattus no	2	AC142069	1.9	246009	2	AC142069
AC136873	Rattus no	2	AC136873	1.9	258901	2	AC136873
AC128167	Rattus no	2	AC128167	1.9	163245	2	AC128167
AC121395	Rattus no	2	AC121395	1.9	213967	2	AC121395
AC128404	Rattus no	2	AC128404	1.9	227451	2	AC128404
AC095816	Rattus no	2	AC095816	1.9	236858	2	AC095816
AC118529	Rattus no	2	AC118529	1.9	324430	2	AC118529
AC117053	Rattus no	2	AC117053	1.9	216261	2	AC117053
AC096021	Rattus no	2	AC096021	1.9	234655	2	AC096021
AC110852	Rattus no	2	AC110852	1.9	265515	2	AC110852
AC129127	Rattus no	2	AC129127	1.9	174010	2	AC129127
AC127118	Rattus no	2	AC127118	1.9	292004	2	AC127118
AC126664	Rattus no	2	AC126664	1.8	231369	2	AC126664
BC009873	Homo sapi	9	BC009873	1.8	1027	9	BC009873
Continuation (4 of				1.8	110000	2	AC114711_3
AC114875	Rattus no	2	AC114875	1.8	240364	2	AC114875
AC093380	Rattus no	2	AC093380	1.8	244216	2	AC093380
AC095250	Rattus no	2	AC095250	1.8	243534	2	AC095250
AC124207	Rattus no	2	AC124207	1.8	256367	2	AC124207
AC121615	Rattus no	2	AC121615	1.8	259169	2	AC121615
AC129703	Rattus no	2	AC129703	1.8	304078	2	AC129703
AC128070	Rattus no	2	AC128070	1.8	314884	2	AC128070
AC112286	Rattus no	2	AC112286	1.8	272490	2	AC112286
AC136569	Rattus no	2	AC136569	1.8	233510	2	AC136569
AC128303	Rattus no	2	AC128303	1.8	243044	2	AC128303
AC096988	Rattus no	2	AC096988	1.8	286491	2	AC096988
BD155096	Primer fo	6	BD155096	1.8	482	6	BD155096
AX875034	Sequence	6	AX875034	1.8	482	6	AX875034
BD158568	Primer fo	6	BD158568	1.8	2148	6	BD158568
AX880750	Sequence	6	AX880750	1.8	2148	6	AX880750
AK023680	Homo sapi	9	AK023680	1.8	2148	9	AK023680
AC097824	Rattus no	2	AC097824	1.8	238052	2	AC097824
AC117040	Rattus no	2	AC117040	1.8	254464	2	AC117040
AC135370	Rattus no	2	AC135370	1.8	208175	2	AC135370
AC110351	Rattus no	2	AC110351	1.8	242324	2	AC110351
AC127927	Rattus no	2	AC127927	1.8	250161	2	AC127927
AC129372	Rattus no	2	AC129372	1.8	173766	2	AC129372
AC122442	Mus muscu	10	AC122442	1.8	180516	10	AC122442
AC079419	Mus muscu	2	AC079419	1.8	196003	2	AC079419
AC133242	Rattus no	2	AC133242	1.8	197101	2	AC133242
AC098921	Rattus no	2	AC098921	1.8	235178	2	AC098921
AC132792	Rattus no	2	AC132792	1.8	245942	2	AC132792
AY007107	Homo sapi	9	AY007107	1.8	1366	9	AY007107
AC090439	Homo sapi	2	AC090439	1.8	199002	2	AC090439
AC097188	Rattus no	2	AC097188	1.8	220314	2	AC097188
AC111351	Rattus no	2	AC111351	1.8	224803	2	AC111351
AC122593	Rattus no	2	AC122593	1.8	252616	2	AC122593
AC096275	Rattus no	2	AC096275	1.8	297409	2	AC096275
AX883047	Rattus no	10	AX883047	1.8	349571	10	AX883047
AC126912	Rattus no	2	AC126912	1.8	171181	2	AC126912
AC140757	Rattus no	2	AC140757	1.8	220229	2	AC140757
AC110855	Rattus no	2	AC110855	1.8	231741	2	AC110855
AC106182	Rattus no	2	AC106182	1.8	235223	2	AC106182
AC133127	Rattus no	2	AC133127	1.8	252324	2	AC133127

Db 421 GTTCCGGCTGCCCTTCCTTCGGGGATCGCAGCGCTGCTCTTCGGAGTTCGCCGCCCTTC 480
Qy 578 CTCTCGACAAAATCCCGGAACTCCGCTCTGCCCGAGCGTCGACACAGTACTGGACCAA 637
Db 481 CTCTCGACAAAATCCCGGAACTCCGCTCTGCCCGAGCGTCGACACAGTACTGGACCAA 540
Qy 638 ATTGCTTTCTCAGCTCTTGGCTGCTCCCTAGCTATTTCCAGAAAGTCTGCTTTGGAG 697
Db 541 ATTGCTTTCTCAGCTCTTGGCTGCTCCCTAGCTATTTCCAGAAAGTCTGCTTTGGAG 600
Qy 698 CCAGCTTTCCGGGGCTGATTCCTACAGATGCTAGATTTTGCSCGAAGTTACAGCGC 757
Db 601 CCAGCTTTCCGGGGCTGATTCCTACAGATGCTAGATTTTGCSCGAAGTTACAGCGC 660
Qy 758 CCTGAGAGCTTCGAGAGCGGAGGAATCTGACGCTCCCAACGCTGACAGAGTCTCTGAG 817
Db 661 CCTGAGAGCTTCGAGAGCGGAGGAATCTGCGCTCCCAACGCTGACAGAGTCTCTGAG 720
Qy 818 TTACATGCGGCTGGACTCTTTCCGGAAGACTCGCGTCTGTCAGTACTCTTGATTTGGCTAG 877
Db 721 TT-CACCTGCGGCTGGACTCTTCGGAAGACTT--GGTTGTCAAGTCTCTTG-ATTGGCTAG 776
Qy 878 AGGAGGACTCCAGTGGCAGTCTGCTCGTCTCAG-CTGGAAAGTTAAACTCAAGCCGAG 936
Db 777 AGGAGGACTCCAGTGGCAGTCTGCTCTCAGACCTGGAGTTAAAACTCAAGCCGAG 836
Qy 937 AAAGAGCTTTAGACTCTGCAGCGCCCACTTTCTCTCGAGCAGCAGCTGTGGGGAGTGG 996
Db 837 AAAGAGCTTTAGACTCTGCAGCGCCCACTTTCTCTCGAGCAGCAGCTGTGGGGAGTGG 896
Qy 997 AGTTGCTGCCAGTAGCTTCAAGCTGGTCTAGTCTCCCAACGAGAACTTTGACTCTTTCAT 1056
Db 897 AGTTGCTGCCAGTAGCTTCAAGCTGGTCTAGTCTCCCAACGAGAACTTTGACTCTTTCAT 956
Qy 1057 CCTCTGGGCTCTGAGCGTTGAGAGCTTAGTAAATTTCAAGTAGTTTCTATCTCTGA 1116
Db 957 CCTCTGGGCTCTGAGCGTTGAGAGCTTAGTAAATTTCAAGTAGTTTCTATCTCTGA 1016
Qy 1117 ACCCTTCTACCTGGACTACCTTCCCGAGTTAGGGCTGCGCTGTCAGAGCAGCGCTGGAG 1176
Db 1017 ACCCTTCTACCTGGACTACCTTCCCGAGTTAGGGCTGCGCTGTCAGAGCAGCGCTGGAG 1076
Qy 1177 GTGGCCAGTTTGTGGTTTCCGAACACTAAACCCAGAGAGCTGCTATCTTTCTGAAGATG 1236
Db 1077 GTGGCCAGTTTGTGGTTTCCGAACACTAAACCCAGAGAGCTGCTATCTTTCTGAAGATG 1136
Qy 1237 GTTGTCAACCTCAGCGCTGGGGCAGAGATGTGGCAACGGCTGAGAGGTGTCCGC 1296
Db 1137 GTTGTCAACCTCAGCGCTGGGGCAGAGATGTGGCAACGGCTGAGAGGTGTCCGC 1196
Qy 1297 CTCTCTACAGAGGCTTCGGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTC 1356
Db 1197 CTCTCTACAGAGGCTTCGGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTC 1256
Qy 1357 TCCAGGCTAAACAAAGGCAAGAGTTACCAACCCCTGACCAAGATAATGGCTATCATAGCC 1416
Db 1257 TCCAGGCTAAACAAAGGCAAGAGTTACCAACCCCTGACCAAGATAATGGCTATCATAGCC 1316
Qy 1417 TGGAGGAGAACATAACCTTCTCGGATGGACCCCAACACATTGACACAGATAACCCAGCAC 1476
Db 1317 TGGAGGAGAACATAACCTTCTCGGATGGACCCCAACACATTGACACAGATAACCCAGCAC 1376
Qy 1477 AGGCGGTGTCCCTGCTGCAGACAGCGCGAGCCCACTGAGAAAAAACAGAAATTTGGTGA 1536
Db 1377 AGGCGGTGTCCCTGCTGCAGACAGCGCGAGCCCACTGAGAAAAAACAGAAATTTGGTGA 1436
Qy 1537 TTCAGAAAGTTTCAACAGAGCCCCCAGGGAAAGCAGTCTGTTTTGTGAATTAACCGGTGAAA 1596
Db 1437 TTCAGAAAGTTTCAACAGAGCCCCCAGGGAAAGCAGTCTGTTTTGTGAATTAACCGGTGAAA 1496
Qy 1597 AAGNATGTGAAGAGCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCCTTC 1656
Db 1497 AAGNATGTGAAGAGCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCCTTC 1556

Qy 1657 CTGTTTTCTACGACACAGTTTGTAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCC 1716
Db 1557 CTGTTTTCTACGACACAG-TTGTAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCC 1615
Qy 1717 CAGTCTACCTTGGNAGCAGCTCTGATTTCTGAAAGTGAAGATTGGGGGAGGAACCTGAGG 1776
Db 1616 CAGTCTACCTTGGNAGCAGCTCTGATTTCTGAAAGTGAAGATTGGGGGAGGAACCTGAGG 1675
Qy 1777 ACCATGGCTTTGATAGCAGTGGCTCCTCTGTCTGAATCAGACGTGGAAACAGGACTCGGAAG 1836
Db 1676 ACGATGGCTTTGATAGCAGTGGCTCCTCTGTCTGAATCAGACGTGGAAACAGGACTCGGAAG 1735
Qy 1837 GCCTTCACTTTTGGAACTCTTTTCCACAGTGTAGATCCTTACAAACCCCAAAACTTTACAG 1896
Db 1736 GCCTTCACTTTTGGAACTCTTTTCCACAGTGTAGATCCTTACAAACCCCAAAACTTTACAG 1795
Qy 1897 CCACGATTCAGACGGCTGCCAGNAATTTGCCCCAGAGACCCATCAGATTCAGGGACATCCT 1956
Db 1796 CCACGATTCAGACGGCTGCCAGNAATTTGCCCCAGAGACCCATCAGATTCAGGGACATCCT 1855
Qy 1957 GGTCTGGCAGCTGTGTGTAGGAGCTGTCCAGGAGGACCCCTTCGGAGACCCCGGACC 2016
Db 1856 GGTCTGGCAGCTGTGTGTAGGAGCTGTCCAGGAGGACCCCTTCGGAGACCCCGGACC 1915
Qy 2017 ATAGTTCCGGGGAGGAAGATGACTGGGAACCCGAGTGCAGATGAAGCAGAGAACTTTAAAT 2076
Db 1916 ATAGTTCCGGGGAGGAAGATGACTGGGAACCCGAGTGCAGATGAAGCAGAGAACTTTAAAT 1975
Qy 2077 TGTGGAACCTTTTCTGTCTCATTTCTGAGACCCCTTACAACTTTTAAATTTTAAAGCTCCTT 2136
Db 1976 TGTGGAACCTTTTCTGTCTCATTTCTGAGGACCCCTTACAACTTTTAAATTTTAAAGCTCCTT 2035
Qy 2137 TTCAAACCGTCAGGAGAAATTTGAAAGGCGTCAGGACTCAAAGGCTCTTCTGAGAGTCA 2196
Db 2036 TTCAAACCGTCAGGAGAAATTTGAAAGGCGTCAGGACTCAAAGGCTCTTCTGAGGCTA 2095
Qy 2197 CAGTGGCTTCTCTGGCCATCATACCTTACTTTTGTAAAGCCAGCTGTGTAGAGAGCC 2256
Db 2096 CAGTGGCTTCTCTGGCCATCATACCTTACTTTTGTAAAGCCAGCTGTGTAGAGAGCC 2155
Qy 2257 AAGAAATTAATTTCTCAGGCTGTGGCTGGGTGAGGCTCTTGTGCGAAGAGATACACC 2316
Db 2156 AAGAAATTAATTTCTCAGGCTGTGGCTGGGTGAGGCTCTTGTGCGAAGAGATACACC 2215
Qy 2317 ATATCAAGAGAAAAAGGTAACTTCTCTGGAAGAGTTACTGAGTATATATAAGTGGTG 2376
Db 2216 ATATCAAGAGAAAAAGGTAACTTCTCTGGAAGAGTTACTGAGTATATATAAGTGGTG 2275
Qy 2377 ATGAGGATCGAAAGGACCATGGGAAGNAATTTGCAAGGGATGGATCGAGTTCAGAAAC 2436
Db 2276 ATGAGGATCGAAAGGACCATGGGAAGNAATTTGCAAGGGATGGATCGAGTTCAGAAAC 2335
Qy 2437 GAATTCAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCAGAGAAAAA 2496
Db 2336 GAATTCAGAAACAGAAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCAGAGAAAAA 2395
Qy 2497 TGTTTAATAGACTGAGGATCGAGTCAAGGACTTACTGTGTGTACAGCAATGTTTAAAGAGT 2556
Db 2396 TGTTTAATAGACTGAGGATCGAGTCAAGGACTTACTGTGTGTACAGCAATGTTTAAAGAGT 2455
Qy 2557 GAAACAGCTGCAACCCGTCGCCACTCTGCTCTTACTTGGAGGTTTCCCTTAAAAA CAAA 2616
Db 2456 GAAACAGCTGCAACCCGTCGCCACTCTGCTCTTACTTGGAGGTTTCCCTTAAAAA CAAA 2515
Qy 2617 CACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAGTT 2676
Db 2516 CACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAGTT 2575
Qy 2677 TGATTAATTTTGGGTAATGTGTCTCAATTAGAAACACAACTCCGATAATGAAGAACTCTCT 2736
Db 2576 TGATTAATTTTGGGTAATGTGTCTCAATTAGAAACACAACTCCGATAATGAAGAACTCTCT 2635

Qy	2737	TATCTGTAATCCTCTCTCTTTCCCTATTTAGTTTGAATGGGGTTTTGTCCTTTTTGAGAGGGT	2796
Db	2636	TATCTGTAATCCTCTCTCTTTCCCTATTTAGTTTGAATGGGGTTTTGTCCTTTTTGAGAGGGT	2695
Qy	2797	CTCACTGCATAATTTCTGTTTGGCCTTAGGTCCTATGTATGTATGATGATAGCAAGGTTGGCCTTGAAGT	2856
Db	2696	CTCACTGCATAATTTCTGTTTGGCCTTAGGTCCTATGTATGTATGATGATAGCAAGGTTGGCCTTGAAGT	2755
Qy	2857	TGGTATTTTCCCTTGCTCTGTTTTCCTGTGCGCACCAATGCCAGCTCGGAAGTGTTTTTAAATA	2916
Db	2756	TGGTATTTTCCCTTGCTCTGTTTTCCTGTGCGCACCAATGCCAGCTCGGAAGTGTTTTTAAATA	2815
Qy	2917	TCCTCTGCTTACCTGGGGTGAGGTTAAATTTTGCACCTTTCAAAACGTTTCTTCTGAGG	2976
Db	2816	TCCTCTGCTTACCTGGGGTGAGGTTAAATTTTGCACCTTTCAAAACGTTTCTTCTGAGG	2875
Qy	2977	CAGGGGCTGGTCTGGTTTTACATACAGGCTTCAGGCCAGCTTAGGGCTATGTGTTGAGACCT	3036
Db	2876	CAGGGGCTGGTCTGGTTTTACATACAGGCTTCAGGCCAGCTTAGGGCTATGTGTTGAGACCT	2935
Qy	3037	AGTCTTAGAGGACAAACAGAAACAAACAGTCAGGTTACTGTGGAACCTGAGGCAGAGGA	3096
Db	2936	AGTCTTAGAGGACAAACAGAAACAAACAGTCAGGTTACTGTGGAACCTGAGGCAGAGGA	2995
Qy	3097	TAGGAAGGTCAGGTATGCCCTGGAACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGA	3156
Db	2996	TAGGAAGGTCAGGTATGCCCTGGAACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGA	3055
Qy	3157	GTTTAGCAAGACCTTTTTACACTCAGAAATGGAAGAGCTGGGATATAGCTCGGTAG	3216
Db	3056	GTTTAGCAAGACCTTTTTACACTCAGAAATGGAAGAGCTGGGATATAGCTCGGTAG	3115
Qy	3217	CAGAGGCCCTGCCCTACATGTGCAAGGTCCTGGGTTCCAGTCCCAGTACTGCAAAATAGAA	3276
Db	3116	CAGAGGCCCTGCCCTACATGTGCAAGGTCCTGGGTTCCAGTCCCAGTACTGCAAAATAGAA	3175
Qy	3277	AGAAAAACATTTGTCTTTGGATAACTATAAGGTTTTAAGCCTCATAGTCAGTCTTAACTCAA	3336
Db	3176	AGAAAAACATTTGTCTTTGGATAACTATAAGGTTTTAAGCCTCATAGTCAGTCTTAACTCAA	3235
Qy	3337	ATTATGATGCAATGAGTCTCTGTGTCCTTTTCTGTTCTTAAATTAATGGGCTTTATGG	3396
Db	3236	ATTATGATGCAATGAGTCTCTGTGTCCTTTTCTGTTCTTAAATTAATGGGCTTTATGG	3295
Qy	3397	GTTTTGTTTTTGATGTTTTTATGCTGTTTATATTTTGGAGTAGGGGATTTAGGCAAGGT	3456
Db	3296	GTTTTGTTTTTGATGTTTTTATGCTGTTTATATTTTGGAGTAGGGGATTTAGGCAAGGT	3355
Qy	3457	CTCACTGGCCTTGAGCTCATGGCAGTTTTTCCCTGCTTCAACTCAGTGTGTTTTGTTTGT	3516
Db	3356	CTCACTGGCCTTGAGCTCATGGCAGTTTTTCCCTGCTTCAACTCAGTGTGTTTTGTTTGT	3415
Qy	3517	TGTGTTTTGTTTTTTTACCTTCATAGATTTGACTTAATGAAGGCAAAAAACCTGTTATCA	3576
Db	3416	TGTGTTTTGTTTTTTTACCTTCATAGATTTGACTTAATGAAGGCAAAAAACCTGTTATCA	3475
Qy	3577	ACCTTAAAGACATTTGATGTCATCTCAGTGGTGGATTTTCTCCCTCTTTTTTTTTTTTCC	3636
Db	3476	ACCTTAAAGACATTTGATGTCATCTCAGTGGTGGATTTTCTCCCTCTTTTTTTTTTTTCC	3534
Qy	3637	CCAGAGCTGAGGACCAAAACCCAGGGCTTTTGGCACTTGCTAGGCAAGCGCTCTACCACTGA	3696
Db	3535	CCAGAGCTGAGGACCAAAACCCAGGGCTTTTGGCACTTGCTAGGCAAGCGCTCTACCACTGA	3594
Qy	3697	GCTAAATCCCAAAACCAACCCCCCTTTTCTTTTTTTTAAAGACATGGTCTTATATAGT	3756
Db	3595	GCTAAATCCCAAAACCAACCCCCCTTTTCTTTTTTTTAAAGACATGGTCTTATATAGT	3654
Qy	3757	CTAGGCTGGCTTTAACTACTATTAGCCAAAGATGGCTTTGAAAGATCCTCCTGTCTCT	3816
Db	3655	CTAGGCTGGCTTTAACTACTATTAGCCAAAGATGGCTTTGAAAGATCCTCCTGTCTCT	3714
Qy	3817	CTTGTTTGGGCTTGACATGCGCATGTGCCATACATCCAGATTTTTTCTCTGTATCTAGTCTCT	3876

3715	Db	 CTTTGTTGGCTTGGCACAATGGGCATGTGCCATATCCAGAAATTTTCTGTATCTAGTGC- 3773
3877	Qy	TTTTTAAGATATTTCCAAAGGACATGCCCAAAAATCCACAGTGCAGTACTTTCTTGACCTGGGA 3936
3774	Db	TTTTTAAGATATTTCCAAAGGACATGCCCAAAAATCCACAGTGCAGTACTTTCTTGACCTGGGA 3833
3937	Qy	AAGCGGGTGTGGTGTCTTTTCAAAGGCAACACCAATTAATGCCAGCTGGTCTGCCAAAC 3996
3834	Db	AAGCGGGTGTGGTGTCTTTTCAAAGGCAACACCAATTAATGCCAGCTGGTCTGCCAAAC 3893
3997	Qy	TTAAACATATTTGGTTTCTGTGCMAACTGTCCATTTTGGAAAGTTTCTGTGTGTGTGTAGTTT 4056
3894	Db	TTAAACATATTTGGTTTCTGTGCMAACTGTCCATTTTGGAAAGTTTCTGTGTGTGTGTAGTTT 3953
4057	Qy	CAGTTCAATGTGGCTCTCTGTGTAGCTTTTCAAGAATGGTAGAAATCATAAAGCACTCTTAA 4116
3954	Db	CAGTTCAATGTGGCTCTCTGTGTAGCTTTTCAAGAATGGTAGAAATCATAAAGCACTCTTAA 4013
4117	Qy	GTAATCAITTCATTTGTAGACATTTTTTTTTTTTTTAACTAGGGGGTATTTGGGAGACATGTC 4176
4014	Db	GTAATCAITTCATTTGTAGACA-TTTTTTTTTTTTAACTAGSGGGTATTTGGGAGACATGTC 4072
4177	Qy	AGATATTCATTTTGTTTTATGTGTCTTAAAAACAGTGTTACTTACACCTGGCTCAGCA 4236
4073	Db	AGATATTCATTTTGTTTTATGTGTCTTAAAAACAGTGTTACTTACACCTGGCTCAGCA 4132
4237	Qy	CTGGCACTTTTCAAACCTGTCTTGGAGACACHTGTAAACTTGGATGGTGAGTTCTCGGTT 4296
4133	Db	CTGGCACTTTTCAAACCTGTCTTGGAGACACHTGTAAACTTGGATGGTGAGTTCTCGGTT 4192
4297	Qy	TTTCATGTGTAAATGTCAACTCAAAGGTCAATATCCAGGTTTGTGTGTCTTACCTTAT 4356
4193	Db	TTTCATGTGTAAATGTCAACTCAAAGGTCAATATCCAGGTTTGTGTGTCTTACCTTAT 4252
4357	Qy	GTAGAGAGAAGCAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGTCTAGTGT-CCCC 4415
4253	Db	GTAGAGAGAAGCAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGTCTAGTGTCCCCC 4312
4416	Qy	CCAAAGCGTCTCTAGATAGTGTGGATCAGGTGTGCTTGGTTTGTTCAGTTAGTACGTC 4475
4313	Db	CCAAAGCGTCTCTAGATAGTGTGGATCAGGTGTGCTTGGTTTGTTCAGTTAGTACGTC 4372
4476	Qy	ACCTGTTGTGCATGGTTGAGGTAATGTGTATCCAGTTCTTTAGTGGTTACATGCACAAA 4535
4373	Db	ACCTGTTGTGCATGGTTGAGGTAATGTGTATCCAGTTCTTTAGTGGTTACATGCACAAA 4432
4536	Qy	GAGAGACCTCTGAGTTCGGGTGTGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCT 4595
4433	Db	GAGAGACCTCTGAGTTCGGGTGTGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCT 4492
4596	Qy	CAGTTTATAATCTCCCTGGTTATTTCCGTTTCAGTGTGATCTAAGTCTGCCTCAGTGTC 4655
4493	Db	CAGTTTATAATCTCCCTGGTTATTTCCGTTTCAGTGTGATCTAAGTCTGCCTCAGTGTC 4552
4656	Qy	ATGATGTTTCATCCACACAAGGTTAGTAAAGAGTGCACACAGAAAACGTTGGTCTTATTT 4715
4553	Db	ATGATGTTTCATCCACACAAGGTTAGTAAAGAGTGCACACAGAAAACGTTGGTCTTATTT 4612
4716	Qy	TTTGAGAACCCCCATTTCTGTGTATTTTATGCACTTCCTTTAGTGAACTCCAGAGTGA 4775
4613	Db	TTTGAGAACCCCCATTTCTGTGTATTTTATGCACTTCCTTTAGTGAACTCCAGAGTGA 4672
4776	Qy	TTAAAGAGTCTGGTTAGTGCCTGGGAATGGGCTAGTTTAGAAGCTATGTTTGGAAAGC 4835
4673	Db	TTAAAGAGTCTGGTTAGTGCCTGGGAATGGGCTAGTTTAGAAGCTATGTTTGGAAAGC 4732
4836	Qy	AGGCAAGTTGACTTTTAGGAAGAAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGG 4895
4733	Db	AGGCAAGTTGACTTTTAGGAAGAAAAGCTGTGACAGTGTGTAGACATTTCTTTTAAACCGG 4792
4896	Qy	ACTGCAAGCTTAAACAACATTTGATTTTCAGATGATTAAGTTTGTGTTCTGAGACCCAGCAC 4955

QY 2512 GGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTAAGAGTGAACAGCCTGCAACC 2571
DB 746 GGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTAAGAGTGAACAGCCTGCAACC 805
QY 2572 CGTGCCCACTCTGTCTCTTACTTGAGAGTGTCCCTTAAAAA CAAACACTGCGACGTGTC 2631
DB 806 CGTGCCCACTCTGTCTCTTACTTGAGAGTGTCCCTTAAAAA CAAACACTGCGACGTGTC 865
QY 2632 TTGGACATGTTTTTAAAGAAACAACTGTATCTAGAGATGACAGTTGATTTATTTTGGGT 2691
DB 866 TTGGACATGTTTTTAAAGAAACAACTGTATCTAGAGATGACAGTTGATTTATTTTGGGT 925
QY 2692 AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGAACTCTTATCTGTAACTCCTCT 2751
DB 926 AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGAACTCTTATCTGTAACTCCTCT 985
QY 2752 CTTTTCTCATTTAGTTGGATGCGGTTTTGTCTTTTTTGGAGGGGTCTCACTGCAATAATTC 2811
DB 986 CTTTTCTCATTTAGTTGGATGCGGTTTTGTCTTTTTTGGAGGGGTCTCACTGCAATAATTC 1045
QY 2812 TGTTTGGCCTAGGCTATGATGTAGACAGGTTGGCCTTGAAAGTTGGTATTTCTTTGCC 2871
DB 1046 TGTTTGGCCTAGGCTATGATGTAGACAGGTTGGCCTTGAAAGTTGGTATTTCTTTGCC 1105
QY 2872 TCTGTTTCTGTGCCACCATGCCAGCTGGAAAGTGTTTTTTAAATATCCTCTGTCTTACCTG 2931
DB 1106 TCTGTTTCTGTGCCACCATGCCAGCTGGAAAGTGTTTTTTAAATATCCTCTGTCTTACCTG 1165
QY 2932 GGGTGAGAGTTAAATTTTTGCACTTTTCAAAACGTTTTCTCTGGAGCAGGGCTGGTCTGG 2991
DB 1166 GGGTGAGAGTTAAATTTTTGCACTTTTCAAAACGTTTTCTCTGGAGCAGGGCTGGTCTGG 1225
QY 2992 TTTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGTGAGACCTAGTCTTTAGAGGACAA 3051
DB 1226 TTTTACATACAGGCTTCAGGCCAGCTAGGGCTATGTGTGAGACCTAGTCTTTAGAGGACAA 1285
QY 3052 ACAGAAACAAACAGTCAGGTTACTGTGGAACCTGAGGCAGGAGGATGAAGGTCACAGGT 3111
DB 1286 ACAGAAACAAACAGTCAGGTTACTGTGGAACCTGAGGCAGGAGGATGAAGGTCACAGGT 1345
QY 3112 ATGCTGGAACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGAGTTTACGACAGCTT 3171
DB 1346 ATGCTGGAACTCAGTGAACTCTGAACTCAAGGCCAGCATGGGAGTTTACGACAGCTT 1405
QY 3172 TTTTACTCAGATGGAAGGAGCTGGGATATAGCTCGGTAGCAGAGGCCCTGCCT 3231
DB 1406 TTTTACTCAGATGGAAGGAGCTGGGATATAGCTCGGTAGCAGAGGCCCTGCCT 1465
QY 3232 ACATGTGCAAGTCTCTGGGTTTCACTCCCGAGTACTGCAAAATAGAAAGAAAAACATTTGTC 3291
DB 1466 ACATGTGCAAGTCTCTGGGTTTCACTCCCGAGTACTGCAAAATAGAAAGAAAAACATTTGTC 1525
QY 3292 TTGGATAACTATAAGGTTTAAAGCTCCTAGTCACTGTTTCACTCAAAATATGATGATGATG 1585
DB 1526 TTGGATAACTATAAGGTTTAAAGCTCCTAGTCACTGTTTCACTCAAAATATGATGATGATG 1645
QY 3352 AGTTCTGTGTGCTTTTCTGTTCTTAAATTAATGAGGCTTTATGGGTTTGTGTTTGTGAT 3411
DB 1586 AGTTCTGTGTGCTTTTCTGTTCTTAAATTAATGAGGCTTTATGGGTTTGTGTTTGTGAT 1645
QY 3412 GTTTTTATGTCGTGTTATTTTGAAGTATGAGGATGAGGAGGCTCTCACTGTGGCCTTG 3471
DB 1646 GTTTTTATGTCGTGTTATTTTGAAGTATGAGGATGAGGAGGCTCTCACTGTGGCCTTG 1705
QY 3472 AGCTCATGGCAGTTTTCTGCTTCAACTGAGTGTGTTTTGTTTTGTTTTGTTTTT 3531
DB 1706 AGCTCATGGCAGTTTTCTGCTTCAACTGAGTGTGTTTTGTTTTGTTTTGTTTTT 1765
QY 3532 TTACTTTTCATAGATTTTGAATTAAGAGGCAAAACCTGTTATCAACCTAAAGACATTTG 3591
DB 1766 TTACTTTTCATAGATTTTGAATTAAGAGGCAAAACCTGTTATCAACCTAAAGACATTTG 1825

QY 3592 ATGTGTCACTTCAGTGGTGGATTTTCTCCCTC-TTTTTTTTTTTTTTCCCCCAGAGCTGAGGA 3650
DB 1826 ATGTGTCACTTCAGTGGTGGATTTTCTCCCTCTTTTTTTTTTTTTTTTTTCCCCCAGAGCTGAGGA 1885
QY 3651 CAAAACCCAGGGCTTTGCACTTGTAGGCAAGGCTCTACCACTGAGCTAAATCCCAAA 3710
DB 1886 CAAAACCCAGGGCTTTGCACTTGTAGGCAAGGCTCTACCACTGAGCTAAATCCCAAA 1945
QY 3711 CCA-CCCCCCCCCTTTTCTTTTTTAAAGACATGGTCTTATATAGTCTAGGCTGGGCTT 3769
DB 1946 CCACCCCCCCCCCTTTTCTTTTTTAAAGACATGGTCTTATATAGTCTAGGCTGGGCTT 2005
QY 3770 AAACCTCAATTACTTAGCCAAAGATGGCTTTGAAAGATCCTCTGTCTCTTTGTTGGCTTG 3829
DB 2006 AAACCTCAATTACTTAGCCAAAGATGGCTTTGAAAGATCCTCTGTCTCTTTGTTGGCTTG 2065
QY 3830 CAGATGCGATGTGCATACATCCAGATTTTTTCTCTGTATCTAGGTCTTTTTTAAGATATTT 3889
DB 2066 CAGATGCGATGTGCATACATCCAGATTTTTTCTCTGTATCTAGGTCTTTTTTAAGATATTT 2125
QY 3890 CCAAGGACTGCCACAAATCCAGATGCACTTCTTCTGACTGGGAAAGCGGGTTGGT 3949
DB 2126 CCAAGGACTGCCACAAATCCAGATGCACTTCTTCTGACTGGGAAAGCGGGTTGGT 2185
QY 3950 GTGCTTTTCAAAGGCACACACCACTTATGCCAGCTGGTCTGCCAAACTTAAACATATTTGG 4009
DB 2186 GTGCTTTTCAAAGGCACACACCACTTATGCCAGCTGGTCTGCCAAACTTAAACATATTTGG 2245
QY 4010 TTTTCTGTGCAAACTGTCCATTTTGGAGGTTTTCTGTGTGTTGTAGTTTTCAGTTGATGTTGG 4069
DB 2246 TTTTCTGTGCAAACTGTCCATTTTGGAGGTTTTCTGTGTGTTGTAGTTTTCAGTTGATGTTGG 2305
QY 4070 CTCTCTGTAGCTTTTCAAGAAATGGTAGAAATCAATAAGCACTCTTAAGTAAATCATTTGCAT 4129
DB 2306 CTCTCTGTAGCTTTTCAAGAAATGGTAGAAATCAATAAGCACTCTTAAGTAAATCATTTGCAT 2365
QY 4130 TGTAGACATTTTTTTTTTTTTTAACTAGGGGGTATTTGGGAGACATGTCCAGATATTTCACTTT 4189
DB 2366 TGTAGACATTTTTTTTTTTTTTAACTAGGGGGTATTTGGGAGACATGTCCAGATATTTCACTTT 2425
QY 4190 TGTTTTATGTCTTAAACACAGTTTACTTACACCTGCTCAGCACTGGCAGCTTTTCA 4249
DB 2426 TGTTTTATGTCTTAAACACAGTTTACTTACACCTGCTCAGCACTGGCAGCTTTTCA 2485
QY 4250 AACCTGTCTTTGGAGACACTGTAACTTGGATGGTGAGTTCTGGGTTTTTCAATGTGTAAAA 4309
DB 2486 AACCTGTCTTTGGAGACACTGTAACTTGGATGGTGAGTTCTGGGTTTTTCAATGTGTAAAA 2545
QY 4310 TGTCAACTACAAAGGTTCAATATATCCAGGTTTTGTTGTGTTTCTACCTTATGTAGAGAGAGCA 4369
DB 2546 TGTCAACTACAAAGGTTCAATATATCCAGGTTTTGTTGTGTTTCTACCTTATGTAGAGAGAGCA 2605
QY 4370 AAGCAGAAAGGCGAGATAGAGCAGCCAGAAAGTGTAGTGTCCCCCCCCAAAGCGTGTTC 4429
DB 2606 AAGCAGAAAGGCGAGATAGAGCAGCCAGAAAGTGTAGTGTCCCCCCCCAAAGCGTGTTC 2665
QY 4430 TAGATAGTGTGGATCAGGTGTCTTGGTTGTTTCAAGTTAGCGTCACTGTGTGCGATG 4489
DB 2666 TAGATAGTGTGGATCAGGTGTCTTGGTTGTTTCAAGTTAGCGTCACTGTGTGCGATG 2725
QY 4490 GTTGAGGTAATGTGTTTACCAAGTCTTTTGTGGTTTACATGCAAAAAGAGAGGAGCTCTGA 4549
DB 2726 GTTGAGGTAATGTGTTTACCAAGTCTTTTGTGGTTTACATGCAAAAAGAGAGGAGCTCTGA 2785
QY 4550 GTCGGGTGTGGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCTCAGTTTATAATCTC 4609
DB 2786 GTCGGGTGTGGGATGAGCTTTCCAGACCTGGCAGGGTAAACTACCTCAGTTTATAATCTC 2845
QY 4610 CCTGGTTATTTCCGTTTGTGATGTATCTTAAGGTCTGCGCTCAGTGTGATGATGTTTCATCCA 4669
DB 2846 CCTGGTTATTTCCGTTTGTGATGTATCTTAAGGTCTGCGCTCAGTGTGATGATGTTTCATCCA 2905
QY 4670 CACACAGGTTAGTAAGAGTGCACGACCAAGAAACGTTGGTCTTTATTTTTTGGAGAACCCCA 4729

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20605

Center clone name: 304_O_9

FEATURES

source

Location/Qualifiers

1..215781
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP23-304O9"
/clone.lib="RPCI-23 Female Mouse BAC"

misc_feature

1..8553
/note="wgs_end_extension"
clone_end:SP6"

repeat_region

complement(2..143)

repeat_region

/rpt_family="URR1A"

repeat_region

441..478

repeat_region

/rpt_family="(TTG)n"

repeat_region

776..815

repeat_region

/rpt_family="(CA)n"

repeat_region

complement(817..1006)

repeat_region

/rpt_family="ID B1"

repeat_region

complement(1092..1181)

repeat_region

/rpt_family="ID3"

repeat_region

1304..1512

repeat_region

/rpt_family="B3"

repeat_region

complement(1754..1928)

repeat_region

/rpt_family="B3"

repeat_region

complement(2578..2681)

repeat_region

/rpt_family="MIR"

repeat_region

2809..2907

repeat_region

/rpt_family="B1_MM"

repeat_region

2902..2957

repeat_region

/rpt_family="B2_Mm1"

repeat_region

complement(2986..3154)

repeat_region

/rpt_family="B3A"

repeat_region

complement(3608..3753)

repeat_region

/rpt_family="B1_MM"

repeat_region

complement(3772..3920)

repeat_region

/rpt_family="B3A"

repeat_region

complement(3974..4059)

repeat_region

/rpt_family="PB1D10"

repeat_region

complement(4101..4207)

repeat_region

/rpt_family="PB1D9"

repeat_region

4828..4971

repeat_region

/rpt_family="B1F"

repeat_region

5328..5474

repeat_region

/rpt_family="B1_MM"

repeat_region

complement(6606..6614)

repeat_region

/note="<30 qual SINGL region"

repeat_region

complement(6658..6664)

repeat_region

/note="<30 qual SINGL region"

repeat_region

complement(6700..6708)

repeat_region

/note="<30 qual SINGL region"

repeat_region

complement(6737..6743)

repeat_region

/note="<30 qual SINGL region"

repeat_region

complement(6823..6844)

repeat_region

/note="<30 qual SINGL region"

repeat_region

complement(6896..6900)

repeat_region

/note="<30 qual SINGL region"

repeat_region

7252..7338

repeat_region

/rpt_family="B1F"

repeat_region

8546..8671

misc_feature
/rpt_family="B1F"
8554..8559
/note="clone_boundary"
clone_end:SP6
site:Boori"
9091..9403
/rpt_family="ORR1D"
9406..9777
/rpt_family="ORR1B-int"
9866..10062
/rpt_family="ORR1D"
10732..10821
/rpt_family="ID B1"
complement(11142..11221)
/rpt_family="PB1D9"
12068..12243
/rpt_family="B2_Mm2"
12341..12406
/rpt_family="ORR1D"
12413..12778
/rpt_family="ORR1C"
complement(13030..13132)
/rpt_family="B1_MM"
14513..14624
/rpt_family="PB1D7"
14642..14699
/rpt_family="G-rich"
14860..14963
/rpt_family="PB1D9"
complement(15187..15305)
/rpt_family="B3"
complement(15492..15644)
/rpt_family="B1_MM"

Query Match 56.7%; Score 3100; DB 10; Length 215781;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3122; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy	2332	AGGTAACCTTCCTGGAGAAGTTACTGAGTATTATATAAGTGTGATGAGATCGCAAAG	2391
Db	156280	AGGTAACCTTCCTGGAGAAGTTACTGAGTATTATATAAGTGTGATGAGATCGCAAAG	156339
Qy	2392	GACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTCCAGAAACGAATTCAGAAACAG	2451
Db	156340	GACCATGGGAAGAAATTTGCAAGGGATGGATGCGAGTTCCAGAAACGAATTCAGAAACAG	156399
Qy	2452	AAGTTGCCATTGGCTACTGCTTGGCCCTTTGAGCAGACAGAGAAATTTTAAATAGACTGA	2511
Db	156400	AAGTTGCCATTGGCTACTGCTTGGCCCTTTGAGCAGACAGAGAAATTTTAAATAGACTGA	156459
Qy	2512	GGATCGAGTCAAAGGACTTACTGTTGACAGCAATGTTAAGAAAGTGAACAGCCTGCAACC	2571
Db	156460	GGATCGAGTCAAAGGACTTACTGTTGACAGCAATGTTAAGAAAGTGAACAGCCTGCAACC	156519
Qy	2572	CGTGCCCACTCTGCTCTTACTTTGAGAGTTTCCCTTAAAAACAAACACTGGCAGCTGTCC	2631
Db	156520	CGTGCCCACTCTGCTCTTACTTTGAGAGTTTCCCTTAAAAACAAACACTGGCAGCTGTCC	156579
Qy	2632	TTGGACATGTTTTTAAAGAAACAACTTGATCTAGAGTGCAGTTCATTTATTTTGGGT	2691
Db	156580	TTGGACATGTTTTTAAAGAAACAACTTGATCTAGAGTGCAGTTCATTTATTTTGGGT	156639
Qy	2692	AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGATCTCTTATCTGTAACTCTCT	2751
Db	156640	AATGTGTCTCATTTAGAAACACCAACTCCGATAATGAAGATCTCTTATCTGTAACTCTCT	156699
Qy	2752	CTTTTCTCTATTTAGTTGGATGTGGGTTTTGTCTTTTTTGAGGGGTCTCACTGCATATTC	2811

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 11068: contig of 11068 bp in length
 * 11069 11168: gap of unknown length
 * 11169 212261: contig of 201093 bp in length.

FEATURES

source Location/Qualifiers

1..212261
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-146P11"

misc_feature

1..1217
 /note="wgs_contig"

ORIGIN

Query Match 44.1%; Score 2412.4; DB 2; Length 212261;
 Best Local Similarity 75.7%; Pred. No. 0;
 Matches 3968; Conservative 0; Mismatches 771; Indels 503; Gaps 59;
 404 CGCTGACGCGCTGAGCTCTGTCTCTCTCTGTGTGAGAACGCCCGCAAGGAAAGGAGAT 463
 Db GCTCTGATGCGCTGAGCTC--CGTACTCTGCGCGTGAAGTTGCCAAGGAAAGGAGAT 140658
 464 GGAGACAGGACGACAGGCGCGGAGCGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 523
 Db GGAGACGCGGACACACAGGCGTC--AAAGCGCGCTACCTCTCGGATGGGCTCTGATTATG 140717
 524 GCTGCGCTTCTCGCGGATCGACGCTGCTCTGCGAGTTCCCGCGCTTCTCTCG 583
 Db CTGCGCTTCTTAGGCGGATCACGCTGCTGCTGAGTTCCGGAACCTTCTTCTCA 140777
 584 ACAAAATCCCGGAACTCCGCTCTGCCCGAGCGTCCGACCAAGTACTGGACCAAAATGCT 643
 Db ACCAAATCTCGAGGACTCTGCTGCTGCGAGGCTCAGACCAAGTACTGGACCAAAATGCA 140837
 644 TTCTCAGCTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
 Db TTCCCGATTCCTTGC---TCTCCCTAGCTTCTCTCAGAGCTGCTTATTTGGAGTCACT 140894
 704 TTCCGGGGCTGATTCCTACAGATGCTAGATTTTCCGCAAGTTACAGCGCCCTGAG 763
 Db TTTGGTGGCAATGATCTGACACAGATGCTATATCTTTACAGAGTTACAGCTCCCTGCG 140954
 764 AGCTTCGAGAGGACGGGAGG--AATCTGACGCTCCACGCGTCCAGAGTCTCTGAGTTACA 822
 Db AGCTTTGAGAGGACGGGAGGAAATCAGCGCTCCAGCTCCAGAAATCTTTGAGTT-CT 141013
 823 CTGCGCTGAGACTCTTCCGGAAGACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 882
 Db CTGCGCTGAGATTTTCTTCGGAAGACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 141073
 883 GGACTCAGTGGCAGTGTCTGCTCTCAGCTGGAAGTTAAACTCAAGGCCAGGAAAGAG 942
 Db GGACTCAGTGGCAGTGTCTGCTCTCAGCTGGAATTTGAACTCAAGGCCAGGAAAGAG 141133
 943 CTTTAGACTCTGACGCGCCACTTTCTCTCTGAGCAGCAGCTGTGGGGAGTGGAGTTGC 1002
 Db CTTTAGACTCTACAGCGCACACTTTCTCTCTGAGCAGCAGCTGTGGGGAGTGGAGTTGC 141193
 1003 TGCCAGTAGGCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062
 Db TGCCAGTAGGCTTCAAAACCGATCTAGTCTCTCCACCGAGAACTTTGCTTCTATCTCTG 141253
 1063 GGCTCTGAGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1122
 Db GGCTCTGAGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 141254
 141254 GGCCTCTGAGCAGCTAGGCTTGGGCAATTTCAAGAGTATTCTATCTCTGCTGAAACCTTT 141313

Qy 1123 CCTACTGGAAGTACCTTTCCCGGAGAGATGTGCGCAACCGCTCGGAGAGGTTGTCGCGC 1182
 Db 141314 CCTACTGGAAGTACCTTTCCCGGAGAGATGTGCGCAACCGCTCGGAGAGGTTGTCGCGC 141373
 Qy 1183 AGTTTGTGGGTTTCCGGAACACTAACCCAGAGAGTGTCTATCTTCTGAAGATGTTGTC 1242
 Db 141374 AACTTGTGGATTTCCGAGCGCTATCCCGAGAGCTGTACCTTTCTGAAGTGTGTTGTC 141433
 Qy 1243 ACCCTCAGCGGTTTCCGCGGAGAGATGTGCGCAACCGCTCGGAGAGGTTGTCGCGCTCTCT 1302
 Db 141434 ACTC-CAGCGGCTGCGTGTGAGATGTGCGCAACCGCTCGCGAGAGTGTCCAGCGCTCT 141492
 Qy 1303 CTACAGAGGCTGCGCGGAAATCCACACTCTGCTGATGAACCGCTAGAAATTTCTCAGG 1362
 Db 141493 CTACAGAGGCTGCGCGGAAATCCACACTCTGCTGATGAACCGCTAGAAATTTCTCAG- 141551
 Qy 1363 CTAACAAAGGCAAGAGTTTACCCACCCCTGACCAAGATATGCTATCATAGCTGCGAGG 1422
 Db 141552 -----AGGCAAGAGTTTACCAACCCCTGACCAAGCAATGGCTATCATAGCTTGGAGG 141603
 Qy 1423 AGGAACATAACCTTTCTCCGGATGGACCCCAACAATTTGCACAGATAACCCAGCAGCGG 1482
 Db 141604 AGGAGCACTGCTTTCTCCGGATGGACCCCGCAACATTTGCACAGATAGGTAGCAGGCTA 141663
 Qy 1483 TGTCCCTGCTGACAGCAGCGCGGAGCCCACTGAGAAAACCCAGAAATTTGGTGAATCAAG 1542
 Db 141664 TGTCCCTGCTGAGGCGAGCCCGCAGCCCACTGAGAAAAGTAGAAATTTGGTGTGAGG 141723
 Qy 1543 AAGTTTACAGAGCGCCCGAGGAGCAGTCTGTTTGTGAATTTACCGTCGGAAGAAAGAT 1602
 Db 141724 AAGTTTACAGAGCGCCCGAGGAGCAGTCTGCTTGTGAGTTAAGTGTGCAAAAAGAT 141783
 Qy 1603 GTGAAGAGGACCACTTAATGCAACTGACCTCTCAGATAGGAGAGAGCTTCTCTCTTTT 1662
 Db 141784 GTGAAGAGGACCACTTAATGTAATTTGACTACTCAGACATAGGGGAAACCTTCTCTTT 141843
 Qy 1663 CTACACAGCAGTTTGTAGCAACAACTGATATATATTTTGGAGGCGCCCGCAGTG 1722
 Db 141844 CTCCAGACCGCTTGTACCAAACTGATAGATTTATTTTGGGAGGAGCTTCCAGTG 141903
 Qy 1723 ACTTGGAGGCGCTGATTTCTGAAGTGAAGTGGGCGAGGACCTGAGGACCATG 1782
 Db 141904 ACTTGGACACGCTCGTTTCTGAAGTGAAGTGGGATGAGGAACCTTGAATATGATG 141963
 Qy 1783 GCTTTGATAGCGATGGCTCCCTGCTGTAATCAGACGCTGGAACAGGACTCCGGAAGGCTTC 1842
 Db 141964 GTTTGATAGTGGCTCCCTGCTGTAATCAGACGGAACAGACTCCGGAAGGCTTC 142023
 Qy 1843 ACCTTTGGAACCTTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTTACAGCCAGA 1902
 Db 142024 ACCTTTGGAACCTTTTCTACAGTGTAGATCTTTACAAACCCCAACACTTTTACAGCCAGA 142083
 Qy 1903 TTACAGCGGCTGCGAGAAATTTGCGGAGAGAGCCATCAGATTCAGGAGACATCTCTGCTG 1962
 Db 142084 TTACAGACTGCTGCGAGAAATTTGCGGAGAGAGCCATCAGATTCAGGAGAGTCTCTGCTG 142143
 Qy 1963 GCAGCTGTGTGTAGGAGCTGTGAGGAGGAGCCCTTCCGAGAGAGCCCGGAGCCATAGTT 2022
 Db 142144 GCACCGGATGTAGGAGGCTCTCAGGCAACACTCTTCCGAGAGCCCTGACAGAGTT 142203
 Qy 2023 CCGGGAGGAGAGTACTGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTTAAATTTGGA 2082
 Db 142204 CTGGGAGGAGAGTACTGGGAACCTAGTGCAGAGGAGCAGAAAATCATAGATTTGGA 142263
 Qy 2083 ACTCTTTCTGTCATTTCTGAGGAGCCCTTCAAACTTTTAAATTTTAAAGGCTCTTTTCAAC 2142
 Db 142264 ACTCTCTGTAATTTCTGAGGAGCCCTTCAAACTTTTAAATTTTAAAGGCTCTTTTCAAA 142323
 Qy 2143 CGTCAGGAGAAATTTGGAAGGCGCTCAGGACTCAAAAGGCTCTTCTGAGGCTCAGAGTGG 2202
 Db 142324 CGTTAGGAGAGAGCTGGAAGGCTGTGAGGAGCTCTGGAAGGAGTCTCTGAGGAGGAG 142383
 Qy 2203 CTTTCTTGGCCATCATACCTTACTTTTGTAAAGGCCCA-----GCTGTTAGAGAGCC 2256

Db 142384 TCTTGTCTGGGCATCAGACCTTACTTTCTTGTGAAGGCA CAAGAACTTAAATTAGAGCC 142443
 QY 2257 AAGAGATAATTTGTCCAGGCTGTGGCTGGGTGGAGCTCTTCTGGAGAGATACACCC 2316
 Db 142444 AAGAAGATAATTTGT -AGGGCTGTGGCTGGGTGGAGCTCTTTTGGTGAAGATACACCC 142502
 QY 2317 ATATCAAGAGAAAAAGGTAACTTCTTGGGAAGATTACTGAGTATTATATAAGTGGTG 2376
 Db 142503 ATATCAAGAGAAAAAGATTAACCTTCTTGAAGAATTACTGAGTATTATATAAGTGGTG 142562
 QY 2377 ATGAGGATGCGAAAGGACCATGGGAAGAAATTTGCAAGGATGGATGCGAGGTTCCAGAAAC 2436
 Db 142563 ATGAGGTCGCAAGGATTAATGGGAAGAAATTTGCAAGGATGGATGCGAAGTCCAGAAAC 142622
 QY 2437 GAAATTCAGAAACAGAGTTGCCATGTGGCTACTGTCTGGCTTTTGAGCACAGAGAAAAA 2496
 Db 142623 GAAATTCAGAAACAGAGTTGGCATTTGGTTACTGTCTTACGCTTTGAGCACAAAGTGTT 142682
 QY 2497 TGTTTAATAGACTGAGGATCGAGTCAAGGACTTACTGTGTACAGCAATGTTAAG--AA 2554
 Db 142683 TAGTAGACTGCAAGACCGT-GTTTCAAGGACTTCTCTGTGTACAGCAGTGTCTAAGCAGA 142741
 QY 2555 GTGAACAGCTGCAACCCGTCGCCACTCTGTCTTACTTTGAGA-GTTTCCCTTAAAAAC 2613
 Db 142742 GGGAGCAGCTGCGAGCCCTCACCCACTCCATCTCTTACTTGAGAGGTTTCTCTTATAGAC 142801
 QY 2614 AAACACTGGCAGCTGCTTGGACATGTTTTAAAGAAACACTTGTTATCTAGAGTCA 2673
 Db 142802 AAACACTGGCAGCTGCTTGGACATGTTTTAAAGAAACGACTTGTTCCTAGTGATGA 142861
 QY 2674 GTTTCATTATTTTGGGTAATGTGTCTCATTTAGAAACAC---CAACTCCGATAATGAAGA 2730
 Db 142862 GTTTAGTTATTTTGGATTAATGTGTCTCATTTAGAAACACACAACTCTAATATGAAGG 142921
 QY 2731 ATCTCTATCTGTAACTCTCTTTTCTTCTAATTTAGTTGGATGTTGGG---TTTTGTCTTTT 2787
 Db 142922 ATCTCTAATCTTTAATCTCTTCTTCTTCTATATAGTGGATGTTGGTTTTTTTTTTTT 142981
 QY 2788 TGAGAGGCTCTACCTGCATAATTCGTTTGGCTAGTCTATGTATGTAGACCAAGTTGG 2847
 Db 142982 TGAGAGGCTCTACCTGCACGCA-----TTGTTGGCTAGGTTCTATGTATGTAGACAAAGTTGG 143035
 QY 2848 CTTTCAAGTTGTGATTTCTTCTGCTCTGTTTCTGTCACCATGCCAGCTGCAAGTGT 2907
 Db 143036 CTTTGAATCGGATCTTTTGGCTTGTCTTCTGTCACCATGCCAGCTGCAAGTGT 143095
 QY 2908 TTTTAAATATCTCTGCTTACCTGGGGTGA-----GAGTTTAAATTTTGCACCTT 2955
 Db 143096 TTTTAAACATCTTCTGTAACTGCGGGTGAGAGGGGCCCTTTTGAGTTTAAATCTGCACTT 143155
 QY 2956 TCARAACTTTTCTTCTGAGGCGAGGCTGTGTTTACATACAGGCTTCAGGCCAGC 3015
 Db 143156 TCARAAATATTTCTTCTGAGGCGAGGCTGTGTCAGG-----CCAG 143195
 QY 3016 TAGGCTATGTGGTGAAGCTTCTTAGAGGACAAACAGAACAAAACAGTCAGGTACT 3075
 Db 143196 CCGGCTATGTGGT-----AGTCTTAGAGGACAAAGCAGAACAAAC--CAGGCTACT 143247
 QY 3076 GTGGAACCTGAGGACAGGAGATGAGAAAGGTCCAGGTATGCTCTGGAACCTCAGTGAACCTGT 3135
 Db 143248 GTGAGACAGAGGCGAGGAGGAGACAGTACAGGCTCTGCTGGAACCTCAG-----GG 143300
 QY 3136 AACTCAAGCCAGCATGGGAGTTTGTAGCAAGCCTTTTACACTCAG-----AATGGAAA 3191
 Db 143301 AGCTCAAGCCAGCGTGGGGCGTTTGTAGTGAACCTTTTGTGAGCTCAGAAATGGAAC 143360
 QY 3192 GGAAGCTGGGATATAGCTCGGTAGCAGAGGCCCTGCTCATGTGCAAGGTCTCTGGGT 3251
 Db 143361 GGAGCTGGGATATAGCTCAGCAGCAGAGGCCCTGCTCATGTG-AAAGCCCTGGG 143419
 QY 3252 TCAGTCCCAGTACTGCAAAATAGAAAGAAAAAATTTGTTCTTGGATTAATCTATAGTTTA 3311

Db 143420 TCAGTCTCCAGTACTGCAAGAAGAGAGAAA---ACAGTCTTGGATAACTATAAGCTTCA 143476
 QY 3312 AGCCTCATAGTCAGTCTTAACTCAAAATATATGATGC---ATTCAGTCTTCTGTG 3361
 Db 143477 AGCCTCATATTCAGTCTCTATCTCAGATTATGCACTATCATCTAAAACTGAGTCTTGTG 143536
 QY 3362 TGCCTTTCTGTCTTAAATTAATGGCTTTATGGGTTT-TGTTTTTGTATGTTTTATG 3420
 Db 143537 TGCCTTTCTATATACTAAATCAATGGGTTTTATGTGTTTGTTCATATATTTGTGTGTC 143596
 QY 3421 TCTGTTATATTTGAGGTAGGGGATTCAGGAGGGTCTCACTGTGGCTTGTAGCTCATGG 3480
 Db 143597 TGTTTTATATTTTGAAGCGGGGATTCAGCAGGGTCTCACTGTGACCTT--GCTTATGG 143654
 QY 3481 CAGTTTTCTGCTTCAACTCAGTGTGTTTTGTTTTGTTGTTGTTTTTACTTTTCA 3540
 Db 143655 CAGTCTTCTGCTTCAACCTCCTGAGTGTGGGTATACCTTGTGTTGTTTTTGGTTGT 143714
 QY 3541 TAGATTTGACTTAATGAAGGCAAAACCTGTTCATACCTTAAAGACATTTGATGTGCAC 3600
 Db 143715 TTGTTTATGTTTT----- 143727
 QY 3601 TTCAAGTGTGATTTTCTCCTCTTTTTTTTTTCCCCAGAGCTGAGGACCAAAACCAG 3660
 Db 143728 ----- 143727
 QY 3661 GGCTTTGCACTTGTCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCCAAACCACCCCCC 3720
 Db 143728 ----- 143727
 QY 3721 CCCTTTTCTTTTAAAGACATGGTCTTATATAGTCTAGGCTGGCTTAAACTCATTTAC 3780
 Db 143728 -----GGTAGTCTAGGCTGGCTTGAATCATTTAC 143757
 QY 3781 TTAGCCAAAGATGGCTTTGAAAGATCTCTCTCTCTCTTGTGTGGCTTGCACATGGCATG 3840
 Db 143758 TTAGCCAAAGATGACTTTGGAGAGATCTCTTGTCTC-----TGGATTGCACATGGCATG 143812
 QY 3841 TGCCATACATCCA-----GATTTTTCTCTATCTAGGCTCTTTTAAAGATATTTCCAAAG 3895
 Db 143813 CACATACATCCAGATTGATTTTTTCTATATCTAGTCTATTTAGCATATTTCCAAAG 143872
 QY 3896 ACTGCCCAAAATCCACAGTGCAGTACTTTCTTGTGACTTGGAAAGCGGGTGTGTCCTT 3955
 Db 143873 ACTGCCCAAAACCACCAATACAGTACTTTCTAGACTTGGAAAGCGCGTGT----TGCTT 143928
 QY 3956 TTCAAGGACACACCATTTATGCCAGCTGCTGTCGCAAACT-----TAAACATATTT 4007
 Db 143929 CTTAAAGGACACGCGAGTGTGCGCCAGCTAGGCTGCCAAACTTTAACATGTAGTATATTT 143988
 QY 4008 GGTTCCTGTGCAAACTGTCCATTTTGAAGGTTTTCTGTGTGTGTGTGTTCAGTGTGAATGT 4067
 Db 143989 GGTTCCTGTGCA---CGCCATTTAGNAGGCACTGTGTGTGTGTGTGTCTCCAGTGTGCTGG 144045
 QY 4068 GGCTCTCTGTAGCTTTCAAGATGGGTGAGAAATCATAAAGCACTCT--TAAAGTAATCATT 4125
 Db 144046 CTTTC---GTAGTCTCAAGATGGGCAAAAGATTGTAAGCAATCTCGGAAGTGTATCATT 144102
 QY 4126 GCATTGTAGACATTTTTTTTT-----TTTAACTAGGGGTATTTTGG 4166
 Db 144103 GCATTGTAGACTTTTGTAGTTTAACTTAGAAGCTTAGAATTTTCTAGGGGGTATCTAG 144162
 QY 4167 GAGACATGTCAGATATTCACCTTTTGTATTATGTCTTAAACCCAGTGTGTACTTACACCC 4226
 Db 144163 GAGACATCTCAGAAA-----TCTGTTTATGTATCTTAAACAGATGTACTTATACCC 144216
 QY 4227 TGCTCTAGCACTGGCACTTTTCAAACTGTCTTGGAGACACTGTAAACTTGTGGATGGTGA 4286
 Db 144217 TGCTCTAGCAGCGGCACTTTTCAAACTGTCTTGGTGGACCTGTAAACTTGGATGGTGA 144276
 QY 4287 GTTCTGGGTTTTTATGTGTAATGTAACATCAAAAGGTCAATATCCAGGTTTGTGTTGT 4346
 Db 144277 GTTCTGAGTTTGAAGTATGAATGTCAAAATGCAAAATACAAAGGTCAA-ATCCAGGTCCATTTGT 144335

QY 4347 TCTACCTTATGTAGAGAGCAAGCAAGAGGCGAGATAGACAGCCAGAAAGTGCTA 4406
Db 144336 TCTACCTTAGATAGAGAGAGCAGACAGAGAA-TACAGATAGAACAGTCTTA 144394
QY 4407 GTGTCCCCCAAGCGCTTCTAGATAGTGTGTGATCAGGTGCTGGTTGGTTGCTCA 4466
Db 144395 GTGTCTCCCGAAGCTTGTCTTTTGAATAGAGTGTGGCTGGGCG---TGCTTGTGCTCA 144451
QY 4467 GTTAGCGCTCAGCTGTGTGATGCTGTGAGGTAATGTGTGTA---CCAGTCTTTTAGTGGTTA 4524
Db 144452 GGTAGCTTCACTCTGTGCATGCGAGCTCTTTAGTCCCAGTCTTTACCTTGGGACTAAA 144511
QY 4525 CATGCACAAAGAGAGACCTCTCAGTCGGGTGTG-GGATGAGCTTTTCCAGACCTG--- 4579
Db 144512 CGTGCATTAAGAGAGCGCTCTGGATTGGTGGTGGAGGAGCTTTTCTCACAAGAACT 144571
QY 4580 GCAGGTAACACTCAGCTTATAAATCTCCCTGGTTATTTCCGTTTGTGATGTGATCTAAG 4639
Db 144572 GGCTGGGTAACTCCTCAGTTTACAACTCTCC--TGGTTATTCTGTCTGATGTGCTCTAAG 144629
QY 4640 GTCTGCCCTCAGTGTG-ATGATGTTTCAACACACACAAAGGTTAGTAAGAGTGCACAGCA 4698
Db 144630 GTCTGCCCTCAGTGTGTCAGTGTATATCCACACACAG---TAGGAGAGTGCAGTACCA 144686
QY 4699 GAAACGTTGGTCTTATTTTGGAGAACCCCAATTTCTGTGTATTTTATGCACTGSCCTTTA 4758
Db 144687 GAAACCTTGGGCTTATTTTGGAGAACCTCATTTCTGTGACTTTTGTGCACCTGCCCTTT 144746
QY 4759 GTGAACCTCC--AGAGTGGATTAAGAGTCTGGTTAGTGGGTGGATGGCTAGTGTTA 4816
Db 144747 GTGAACCTCCAGAGAGGGGTTGCACAGCTCTGGTTAGTGTATGGAATGGGCCACAGAA 144806
QY 4817 GAAGCTATGTTTGAAGACGAGCAAGTTGACTTTAGAGAAAGAGCTGCACAGTGTGTA 4876
Db 144807 GCTGTGTTGTTGGAAGAGCGGCTGTGTGACTTTAGGAAGAAACTGTGCAATGCGTA 144866
QY 4877 GACATTTCTTTT-AAACCGGACTCAGCTTAACACACACTTGAATTTTCAGATGATTAGGTTT 4935
Db 144867 GACATTTCTTTTAAACCCAGAGTGCAGCTTAACACTGATTTTCAGAGGATTAGTGTGTTT 144926
QY 4936 TTGTTTCTGAGACCCAGCACTGTATA----- 4962
Db 144927 TGTTTCTGAGACCTAGCACCTGAAATTTTGCAAGGACGGGATATCTATACATTGACGA 144986
QY 4963 -----TTTAAATCTGTCAGATTAACCTTCACTATCAATGAGTAATG 5009
Db 144987 TGTGTTGAGATGTTTGAATATCGTTCAGTTTACACTTCACTGTCAATGAAATAATG 145046
QY 5010 ACTCATGCTGCGACACATGTC-----CTGATGGTGGCAAGAACAGAGGATCT----- 5057
Db 145047 ACTGTTAAATATACCATGCTGCGAACAATGTTGGCAAGACAGAGGGTTTATATA 145106
QY 5058 --TTGACTGAAGAGAGAAA-----ACTGTCTATGTCTATCCAGGCCCCAGAGAAAGAA 5107
Db 145107 AATTGACTGAAGAGAGAGAAAACAAAGTTACTGTCTATGTCTATCCCAATCCAG-GGAAGAA 145165
QY 5108 CACCTCCAGCGCAGCGCAGGCGAGCGATGTTGTTCTAGTTGATATACACATTCA 5167
Db 145166 CACCTCAAGGCAAGCGCAGGCGATGTTGTTCTAGCTG-----AATACTTTACACATGCA 145218
QY 5168 AGTCTTGAGTGGTGTCTTT--AGATCTGTGTAGCATGTGAGGCTCTGTACAGGTGGGGG- 5224
Db 145219 GGTCTTACAGTGTGCTTTAGAGATCTGTGTAGCATGTGAGGCTCTGTACAGGTGGGGAC 145278
QY 5225 CCACACTTCTGAGGGCTGAATATGGGCAACCCCTTTATCTAATCTGAATCAAAACCGTCA 5284
Db 145279 CCACACTTCTGAGTGTGAATATGGGCAAC-----CTAATCTGAATCAAAACCGTCA 145332
QY 5285 AATTTTATTTTATATTTAAGAAA-----GAGT 5314
Db 145333 GATTTTATTTTATATTTAAGAAAAGTAGAGATGGAATGAGTTGTTGTTTTTTTTTTT 145392

QY 5315 TGGGGAATGACATTTTTTTGAGTTGGCTTTTCAGCTCAGTCATCTTTACGTGTAAC--GTG 5372
Db 145393 TTTGGAGTCAGATTTTTTTGAGTTGGC---TTTCAGATTAGTCATTTTACATGTAACAAATG 145449
QY 5373 GAGATTTGATAGCTCAGATTTATTTGTATATAATTTATTAACATACTGTAAATTTGTAAT 5432
Db 145450 TAGGCTTTATAGCTCAGATTTACATTTGTATATAATTTATTAACATACTGTAAATTTGTAAT 145509
QY 5433 AATATATTTTGCATTTTAA 5454
Db 145510 AATATGTTTGCATCATTAA 145531

RESULT 5
AC094236/c
LOCUS
DEFINITION
AC094236 240836 bp DNA linear HTG 14-NOV-2002
Rattus norvegicus clone CH230-3B24, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
AC094236
VERSION
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 240836)
Muzny, D., Marle, H., Metzker, M., Lee, S., Amin, A., Angiano, D.,
Allen, C., Allen, H., Alebrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Dengson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dungan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Cantar, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, C.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowik, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhera, L., Loulaeged, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Plopper, F., Paul, H., Perez, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Soderstrom, E., Song, X.-Z., Sorelle, R., Soles, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 240836)
Worley, K. C.

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

Rat Genome Sequencing Consortium.
3 (bases 1 to 240836)
Direct Submission

Submitted (14-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 14, 2002 this sequence version replaced gi:22772949.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAGP

Center clone name: CH230-3B24

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 194130 bases at least Q40

Consensus quality: 198891 bases at least Q30

Consensus quality: 202078 bases at least Q20

Estimated insert size: 194853; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 184581: contig of 184581 bp in length
* 184582 184681: gap of unknown length
* 184682 238285: contig of 53604 bp in length
* 238286 238385: gap of unknown length
* 238386 239477: contig of 1092 bp in length
* 239478 239577: gap of unknown length
* 239578 240836: contig of 1259 bp in length.

FEATURES
source

1. 240836
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-3B24"
1984..3036
/note="wgs_contig"

misc_feature

ORIGIN

Query Match	40.0%	Score 2188.8;	DB 2;	Length 240836;
Best Local Similarity	75.3%	Pred. No. 0;		
Matches 3655;	Conservative	0;	Mismatches 704;	Indels 495; Gaps 55;

Qy	791	CGCTCCACCGTGCAGAAAGTCTCTGAGTTACATCGCGCTGGACTCTTCGGGAAGACTCG	850
Db	174258	CGCTCCACCGTGCAGAAATCTTTTGGTT-CTCTGCGGCTGGATTTTCTTCGGAAGACT	174200
Qy	851	CGTCTGTCAGTACTCTTTCGATTGGCTAGAGGAGGACTCCAGTGGCAGTGTCTGTCCTCAG	910
Db	174199	TGGTGTGTCAGTCTCTTGTATCGGCTAGAGGAGGACTCCAGTGGCAGTGTCTGTCCTCAG	174140
Qy	911	ACTGGAAGTTAAACTCAAGGCCAGGAAAGAGCTTTTAGACTCTGACGCGCCACTTTTCCT	970
Db	174139	ACCTGGAATTTGAATCAAGGCCAGGAAAGAGCTTTTAGACTCTTACAGCGCACACTTTCT	174080
Qy	971	CCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTTCAAGCTGTCTAGT	1030
Db	174079	CATGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCGCTCTCTTCAAAACCGACTCTAGT	174020
Qy	1031	CTCCACCGGAGAACTTGACTCTTTCATCTCTGGGCTCTCTGAGCGTTCAGAGCTTTAGGTAA	1090
Db	174019	CTCCACCAAGAACTTGGCTCTTTCATCTCTGGGCTCTCTGAGCACTCAGAGCTTTGGCAA	173960
Qy	1091	TTTCAAGGTAGTTTCTCTATCTCTGAAACCTTCTTCTCTGAGTACCTTCCCGAGTTAGG	1150
Db	173959	TTTCAAGGTAGTTTCTCTATCTCTGAAACCTTCTTCTCTGAGTACCTTCCCGAGTTAGG	173900
Qy	1151	GCTGCGCTGTGAGAGCGCTGAGGTGGCCAGTTTGTGGTTTCCGAACTTAACCCC	1210
Db	173899	GCTGCGCTGTGAGAAACAGCATCTGAGTGGCCAACTTTGTGGATTTCCGAGCGCTATCCCC	173840
Qy	1211	AGAGAGCTGTCTCTTCTTGAAGATGGTTGTCAACCTCTCAGCGTTCGGGCGAGAGTGTCT	1270
Db	173839	AGAGAGCTGTCTCTTCTTGAAGTGGTTGTCTCACTC-CAGCCGCTGGTGTGAGTGTCT	173781
Qy	1271	GGCAACCGCTGAGAGAGTGTCCGCTCTCTCTACAGAAAGGCTTCCGGAATCCACCA	1330
Db	173780	GGCAACTGCTGGCGGAGTGTCCAGCCCTCTCTACAGAAAGGCTTCCGGAATCCACCA	173721
Qy	1331	CCTTGTATGAAACGGGTAGAAATTCCTCAGGGCTTAACAAGGGCAGAGTTACCCACCC	1390
Db	173720	CATTGCGATGAAACGGGTAGAAATTCCTTTCAG-----AGCAAGAGTTTACCAACCC	173670
Qy	1391	TGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAACATAAAGTCTTCCGGATGGACCC	1450
Db	173669	TGACCAAGACATGGCTATCATAGCTTGGAGGAGGAGCACTGCTTCTCCGGATGGACCC	173610
Qy	1451	ACAACATTGCAGATAAACCCAGCAGAGCGGTGTCCCTCTGCTGCGAGACAGCGCGGAGCC	1510
Db	173609	GCAACATTGCAGATAAAGGTAGCAGAGGCTATGTCCCTCTGCTGGAGCCAGCCCGCAGCC	173550
Qy	1511	CAGTGAAGAAACCCAGAAATGGTGATTCAAGAGTTTCAAGAGCTTCCAGAGCCCGGAGAGCAG	1570
Db	173549	CAGTGAAGAAACCCAGAAATGGTGATTGAGAGTTTCAAGAGCTTCCAGAGCCCGGAGAGCAG	173490
Qy	1571	TCTGTTTTGTGAATTAACCGTGGAAAGAAATGTGAAGAGGACCACTAATCAACTGA	1630
Db	173489	TCTGTTTTGTGAATTAAGTGTGCAAGAAAGATGTGAAGAGGACCACTAATCAACTGA	173430
Qy	1631	CCTCTCAGATPAGAGGAGAGAGCTTCTCTGTCTTCTACAGACCAAGTTGTGTAACCAAACT	1690
Db	173429	CTACTCAGACATAGGGGGAACCTTCTCTGTCTTCTCCAGACCAAGTTGTGTAACCAAACT	173370
Qy	1691	GATAGATTATTTTGGAGGGCCCCCAGTGTGATTTGGAAGCCAGCTCTGATTTCTGAAG	1750
Db	173369	GATAGATTATTTTGGAGGGAGCCCTCCAGTGTGATTTGGAAGCCAGCTCTGATTTCTGAAG	173310
Qy	1751	TGAGGATTTGGGGCGAGGAACCTCAGGACGATGGCTTTTGTATGAGGATGGCTCCCTGTCTGA	1810
Db	173309	TGAGGATTTGGGATGAGGAACCTCAGGATGATGATGTTTTGTATGATGAGGCTCCCTGTCTGA	173250
Qy	1811	ATCAGACGTGGAAACAGGACTTCGGAAGGCGCTTACCTTTTGGAACTCTTTTCCACAGTGTAGA	1870

Qy	3984	TGGTCTGCAAACT-----TAAACAATTTGGTTTCTGTCAAACTGTCCATTGGAA	4035
Db	171284	TAGGCTGCAAACTTTAAACATGTAGTATATTTGGTTCTGTGCAA--CGCCATTTAGAA	171288
Qy	4036	GGTTTCTGTGTTCTAGTTTCAGTTTGAATGTGGCTCTCTGTAGCTTTCAAGATGGGTA	4095
Db	171227	GGACCTCTGTGTTGTAGTCCAGTTGACTGGCTTTC---GTAGTCTCAAGATGGGCA	171171
Qy	4096	GAAATCATAAAGCACTCT--TAAGTAATCATTTGCCATTTAGACATTTTTTTTTT-----	4146
Db	171170	AAGATTGTAAAGCATCTGGGAAGTGATTCATTTGCATTTGAGACTTTTAGTTTAAACTT	171111
Qy	4147	-----TTTAACTAGGGGGTATTTGGGAGACATGTCCAGATATTCATTTTGTGTTT	4194
Db	171110	AGAAGCCTAGAAATTTTTCTAGGGGGTATCTAGGAGACATCTCAGAAA-----TCTGTITT	171057
Qy	4195	TATGTGCTTAAACACAGTGTACTTACACCTGCTCAGCACTGGCACTTTTCAAACTT	4254
Db	171056	TATGTATCTTAAACACAGTGTACTTATACCTGCTCAGCACGGGCACTTTTCAAACT	170997
Qy	4255	GTCITGGAGACACTGTAAACTTGGATGTGCAAGTTCTGGTTTTCATGTGTAATAATGCA	4314
Db	170996	GTCITGGTGACCTGTAACTTGGATTTGCAGTTCTGAGTTTGAAGTATGAATGCA	170937
Qy	4315	ACTACAAAGGCTCAATATCCAGTTTTGTTGTTTACCTTATGTAGAGAAAGCAAGCA	4374
Db	170936	AATACAAAGGTCAA--ATCCAGGTCCATTTGTGTTCTACCTTAGATAGAGAAAGCAGAGCA	170878
Qy	4375	GAAGGGCAGATAGACAGCAGCAAGAGTACTAGTGTCCCCCAAGCGTCTCTAGAT	4434
Db	170877	GAAT--TACAGATAGAACAGTNCANAGTCTTAGTGTCTCCCGAAGCTGTCTTTGAT	170819
Qy	4435	AGTGTGGGATCAGGTGTGCTTGGTTTTCAGTTTAGCGTCACTGTTGTGCATGTTGA	4494
Db	170818	AGAGTGTGGCTTGGCG--TGCTTGTTCAGGTAGCTTCACTCTGTGCATGCGACG	170762
Qy	4495	GGTAATGTGTTACCAAGTTCTTTTAGTGGTTACA--TGACAAAAGAGAGACCTCTGAGTC	4552
Db	170761	TCITTAGTCCCAGTCTTTACCTCGGACATAAACGTCGCACTAAGAGAGCGCTCTCGAAT	170702
Qy	4553	GGGTGTG--GGATGAGCTTTCCAGACCTG---GCAGGGTAAACTACCTCAGTTTATAATC	4607
Db	170701	GGTCGTGAGGATGAGCTTTTCTCACAGAACTGGCTGGTAACTACTCAGTTTACAATC	170642
Qy	4608	TCCCTGGTTATTTCCGTTTGATGTGATTAAGGCTCGCTCAGTGGTG--ATGATGTTCAAT	4666
Db	170641	TCC--TGGTTATCTGTCTGATGTCTCTAAGTCTGCTCTCAAGTGTGTCAGTGTGTTAAT	170584
Qy	4667	CCACACACAGGTTAGTAAAGGTGCACACCAAGAACGTTTGGTCTTATTTTTTGAAACCC	4726
Db	170583	CCACACACA--GTAGGAGGAGTGCAGTACCAAGAACCTTGGGCTTATTTTTCGAAGACCC	170527
Qy	4727	CCATTTCTGTGTATTTTATGCACCTGCTCTTGTAGTGAACTCC--AGAGTGCATTAAGAGT	4784
Db	170526	TCAITTTCTGTGACTTTGTGTGCACCTGCTTTTGTGAACTCCAGAGAGGGCGTGCACAGT	170467
Qy	4785	CTGTTTATGTCCTGGGAATGGGCTAGTTTGAAGCTATGTTTGGAAAGCAGGCAAGTT	4844
Db	170466	CTGTTTATGCTATGGAAATGGGCCACAGAAGCTGTGTTGTTGGAAGCAGGCGTGT	170407
Qy	4845	GACTTTAGGAAGAAAGCTGTGCAGTGTGTAGACATTTCTTTTT--AAACCGGACTGCAGC	4903
Db	170406	GACTTTTAGGAAGAAAACTGTGACAAATCGGTAGACATTTCTTTTAAAAACCAAGTGCAGC	170347
Qy	4904	TTAACAAACACTTGATTTTTCAGATGATTAGGTTTGTGTTTCTGAGACCCAGCACTGTATA-	4962
Db	170346	TTAACACTGATTTTCAGAGGATTAGTGTGTTTGTGTTCTGAGACCTAGCACTGAATAT	170287
Qy	4963	-----TTTTAAAAATTTGTTCC	4977
Db	170286	TTGCAAGGCGAGGATATCTATACATTCAGCATGTTGTTGAGATGTTTCAAAATCGTTCC	170227

QY	4978	AGATTACACCTTCACTATCAATGATGTAATGACTCATGCTGCAGACATGTC-----	5030
DB	170226	AGTTTACACCTTCACTGTCAATGAATTAATGACTGGTTAAATAATCAACCTGCCTGCAGAA	170167
QY	5031	CTGATGGTGGCAAGAACAGAAAGGATCT-----TTGACTGAAGCAGAGAAA-----A	5075
DB	170166	CAAATGGTGGCAAGACAGAGAGGGTTTTTATAAATTGACTGAAGCAGAGAAAACAAGTTA	170107
QY	5076	CTGTCAATTGTGTCATCCGAGCCCCCAGGAAAAGAACACCTCCAAGGCCAGGCCGAGGC	5135
DB	170106	CTGTCAATTGTGTCAT--CCATCCACGGGAAGAACAACCTCAAGGCAAGCGCAGGCATGCGGT	170048
QY	5136	AGGCATGGTGGTCTCTAGTTGAATACACATTCAAAGTCCTTGCAGTGGTGCTTT--AGATCTG	5193
DB	170047	TCTAGCTG-----AATACTTTACATGACGAGTCTTACAGTGGTGTCTTATAGAGATCTG	169995
QY	5194	TGTAGCATGTGAGGCTCTGTACAGGTGGG--GCCACACTTCTGAGGGCTGAAATGTGGCA	5252
DB	169994	TGTAGCATGTGAGGCTCTGTACAGGTGGGGCCCCACACTTCTGAGTGTGAAATGTGGCA	169935
QY	5253	ACCCTTTATCTAACTTGAAATCAAAACCGTCGAAATTTTATTTTATTAATTTAAGAAA--	5310
DB	169934	ACC-----CTAACCTTGAATCAAAAACCGTCAGATTTTATTTTATTAATTTAAGAAAAG	169881
QY	5311	-----GAGTTGGGGAATGACATTTTTTTGAGTTGGCCT	5342
DB	169880	TAGAGATGGAATGAGTTTGTGTTTTTTTTTTTTTTTGGACTGAGATTTTTTGGAGTTGGC--	169823
QY	5343	TTTCAGCTCAGTCATTTTACGTGTAAC--GTGGAGATTTGATAGCTCAGATTTATTTGCT	5400
DB	169822	TTTCAGATAGTCATTTTACATGTAACCTAATGTAGGCTTTATAGCTCAGATTTACATTGT	169764
QY	5401	ATATAATTTATTAACATCTCTAAATGTTAATAATATTTTGGCAATTTATTA	5454
DB	169763	ATATAATTTATTAACATCTCTAAATGTTAATAATATTTTGGCAATTTATTA	169710

RESULT 6
AC128396/c
LOCUS
DEFINITION
AC128396 200490 bp DNA linear HTG 26-SEP-2002
Rattus norvegicus clone CH230-194B23, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
AC128396
AC128396.4 GI:23232373
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 200490)
Muzny,D,Maritz, Metzker,M,Lee,, Abramzon,S,, Adams,C,, Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceszar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Rume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshehwa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeil, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Narkervic, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakeleneh, O., Okwunu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, P., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojao, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smaja, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villagana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 200490)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 200490)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 26, 2002 this sequence version replaced gi:22771040.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYZL
Center clone name: CH230-194B23
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 131649 bases at least Q40
Consensus quality: 136220 bases at least Q30
Consensus quality: 138894 bases at least Q20
Estimated insert size: 161200; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.
* 1 21048: contig of 21048 bp in length
* 21049 108746: gap of unknown length
* 21149 108746: contig of 87598 bp in length
* 108747 108846: gap of unknown length
* 108847 161819: contig of 52973 bp in length
* 161820 161919: gap of unknown length
* 161920 176640: contig of 14721 bp in length
* 176641 176641: gap of unknown length
* 176741 187347: contig of 10607 bp in length
* 187348 187447: gap of unknown length
* 187448 200490: contig of 13043 bp in length.
FEATURES
source
1..200490
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-194B23"
1..1226
/notes="wgs contig"
misc_feature
13967..15951
/notes="wgs contig"
misc_feature
19893..21048
/notes="wgs contig"
misc_feature
21149..23404
/notes="wgs contig"
misc_feature
29744..33119
/notes="wgs contig"
misc_feature
48098..50488
/notes="wgs contig"
misc_feature
71079..72931
/notes="wgs contig"
misc_feature
72982..74735
/notes="wgs contig"
misc_feature
89602..91141
/notes="wgs contig"
misc_feature
176741..177999
/notes="wgs contig"
misc_feature
185991..187347
/notes="wgs contig"
misc_feature
187448..188936
/notes="wgs contig"
ORIGIN
Query Match 30.9%; Score 1689.2; DB 2; Length 200490;
Best Local Similarity 84.5%; Pred. No. 0;
Matches 2022; Conservative 0; Mismatches 323; Indels 49; Gaps 9;
QY 1 CGGTCTCTCGCTTCGCGCTTCGAGCTTCGCGGTGCGGCTGCGGCCATTTTGAGCTTCGC 60
DB 197561 CGGTCTCTCGCTTCGCTTCGCTTCGAGCTTCGCGGTGCGGCTGCGGCCATTTTGAGCTTCGC 197502
QY 61 TTCTTTGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCG 120
DB 197501 TTCTTTGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCG 197446
QY 121 TGGGCGCCACCGGAAACGCGCGCTGCTCTCGCTCGCGCGCGCGCGAGGAGGCTCTTCT 180
DB 197445 TGTGGGCGCGC-----GGTCTCGGAAACGTTGCGCGCGCGCGCTGTGAGGGGTCTTCT 197392
QY 181 CTATGTTGAGCGGATCTCACACGCGCTTAGAGCGTCTCTTCTTCCCTAGCGGATGGACCTA 240
DB 197391 CTATGTTGAGCGGATCTCACACGCGCTTAGAGCGCTTCAT-TCTCGCGAGATGGAGCTA 197333
QY 241 ACCGCGTTCGCGCGCTTCGCGCGGCTTCGCGCGCTTCGCGCTTCGCGCTTCGCGCGGA 300
DB 197332 GTCGCGCTTCGCGCGCTTCGCGCGGCTTCGCGCGCTTCGCGCTTCGCGCTTCGCGCGGA 197273
QY 301 AGCGCGCTTCGCGCGCTTCGCGCGGCTTCGCGCGGCTTCGCGCTTCGCGCTTCGCGCTTCG 360
DB 197272 AGCGCGCTTCGCGCGCTTCGCGCGGCTTCGCGCGGCTTCGCGCTTCGCGCTTCGCGCTTCG 197216

Qy 301 AGCGCGCTCTCTGGGCTCTCTGCGGCGCGGGAATCGAGTCGAGTACCACTCCG 360
 Db 152764 AGCGCGCTCTCTGGG---CTCTGCTGGCGCGGGAATCGAATGCACTACCACTCCG 152708
 Qy 361 TGGCTGGGCAAGCGGAGACTGTGTAGACCTCGGATCCAGCTTGGCTGTGAGCGCGCTGAG 420
 Db 152707 TGTCTGGGCAAGGCTTAAGACTGTGTAGGCTTCGGAATCTGCGCGCTCTGAGCGCGCTGAG 152648
 Qy 421 CTCTGTCTCTCTCTCTGCTGAGAGCGCCCAAGGAAGAGATGGAGACAGGAACGACCA 480
 Db 152647 CTC---CGTACTGTGCTCCGAGAGTGTCCCAAGGAAGAGATGGAGHCGGGAACACACA 152591
 Qy 481 GGGCGCGAAGCGGCTGGGCTCTGGCTGGGCTCTCTGGTTCGGGCTGCCCTTCCTTCGGC 540
 Db 152590 GGGCTGGGAAGCGGCTACCTCTCGGATGGGCTCTCTGGTTCGGGCTGCCCTTCCTTCGGC 152531
 Qy 541 GATCGCAGCTGTCTTTCGGAGTTCGGCGCTGCCGCTCTCTCTCGACAAATCCCGGAACT 600
 Db 152530 GTTTCGACGCTGCATTTCCGAGTTCGGAAGGCTTCTTCTCGACCAATCTCGAGGTCT 152471
 Qy 601 CCGCTCTGCCGAGCGTTCGACACAGGTACTGACCAAAATGCTTTCTCAGCTCTCTTGCCC 660
 Db 152470 CCGCGCGCCGAGCTTCGACACAGGTACTGACCAAAATGCTTTCCAGCTCTCTTGCCC 152411
 Qy 661 TGCTCCCTAGCTTATCCAGAAGCTGTCTGCTTTGGAGCAGCTTTCCGGGGGCTGATTC 720
 Db 152410 TGCTCCCTAGCTTGCTCCAGAAGCTGCTTATTTGGAGTCAGCTTTTGGTGGCATGATTC 152351
 Qy 721 CTACAGATGGCTAGATTTTCCGGAAGTTACAGGCGCTGAGAGCTTCGAGAGGACGGG 780
 Db 152350 CGACCAAGTGGTATGATTTTTCAGCAAGTTACAGCGCTCTAAGAGCTTTGAGAGGACGGG 152291
 Qy 781 AGGAATCTGACGCTCCCAAGTGCAGAAAGTCTCTGAGTTACACTCGGCTGCACTCTTCG 840
 Db 152290 AGGAATCGCGCTCCCAAGTGCAGAAATCTTTGAATTCCTCTGCGCTGCACTCTTCT 152232
 Qy 841 CGAAGACTCGCGTCTCAGTACTCTTGCAATGGCTAGAGGAGGACTCCAGTGGCAGTGC 900
 Db 152231 TCGGAAGACTTGGTGTGTCAGTTCTCTTGATTTGGCTAGAGGAGGACTCCAGGGGCACTGC 152172
 Qy 901 TCGTCTCTCAGA-CTGGAAGTTAAACTCAAGGCCCAAGGAAGCTTTAGACTCTGCGAGC 959
 Db 152171 TCGTCTCTCAGACCTGGAAATCGAAATTCAGGCCCAAGGAAGAGCTTTAGACTCTCAGAGC 152112
 Qy 960 CCACTTTCTCTCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCTTCAA 1019
 Db 152111 CACACTTTCTCTTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCGTAGTCTTCAA 152052
 Qy 1020 GCTGGTCTAGTCTCCCAAGAGAACTTGAATCTTCACTCTGCGGCTCTGAGCGTTGAG 1079
 Db 152051 ACCGCTCTAGTCTCCCAAGAGAACTTGGCTCTTCACTCTGCGGCTCTGAGTACTCAG 151992
 Qy 1080 AGCTTAGGTAATTTCAAGGTAGTTTCTATCTCTGAACTCTTCTACCTGGACTACCTT 1139
 Db 151991 AGCTTGGGCAATTTCAAGGTAGTTTCTATCTCTGAACTCTTCTACCTGGACTGCTCTT 151932
 Qy 1140 CCCCAGTTAGGCTGCGCTGTGAGAGCAGCTGTGAGGTGCGCACTTTGCGGTTTCCGA 1199
 Db 151931 CCCCAGTTAGGCTGCGCTGTGAGACAGCTGTGAGGTGCGCACTTTGCGATTTCCGA 151872
 Qy 1200 ACATAACCCAGAGAGCTGTATCTTTCTGAAGATGGTTGTCACTCTGAGCGGTGCGG 1259
 Db 151871 ACATACTCCAGAGAGCTGTACTCTTTCTGAAGATGCTTGTCACTCTGAGCGCTGTGT 151812
 Qy 1260 GCAGAGATGTCGGCAACGCTGAGAGAGGTGTCGCTCTCTCTACAGAGGCTGCGG 1319
 Db 151811 GCTGAGATGTCGGCAACTGCTGCGGAGGTGTCGCGCTCTCTCTACAGAGGCTGCGCA 151752
 Qy 1320 GAAATCCACCACTTCTGATGAACCGCTAGAAATTTCTTCCAGGCTTAACAAAGGCAAGAG 1379
 Db 151751 GAAATCCACCACTTCTGATGAACCGCTAGAAATTTCTTCCAGGCAACAAAGGCAAGAG 151692
 Qy 1380 TTACCCACCCCTGACCAAGATAATGGGCTATCATAGCTTGGAGGAGGAACATAACCTTCTC 1439

Db 151691 TTACCCACCCCTGACCAAGACAATGGCTATCATAGCTTGGAGGAGGACATGCTTCTC 151632
 Qy 1440 CGGATGGAACCAACAATTCACAGATAACCCAGCACAGCGGTGTCCTCTGCTGCAGAC 1499
 Db 151631 CGGATGGAATCCGAAACAATTCACAGATAAGGACAGCAGTCTCCCTCTCTGGAGCC 151572
 Qy 1500 AGGCGGAGCGCCACTGAGAAAAAACCAAGAAATGGTGATTTCAAGAAAGTTTCACAGAGCCCC 1559
 Db 151571 AGCCGCGAGCCACTGAGAAAAAAGTAGAATTTGGTGTTCAGGAAGTTTACAGAGCCCC 151512
 Qy 1560 CAGGGAAGCAGTCTGTTTTGTGAATTTACCCGTGGAAGAAAAAATGTTGAAGAGGACCACT 1619
 Db 151511 CAGGGAAGCAGTCTGTTCTGTGATTTAAGTGTGCAAAAGAAATGTGAA----- 151464
 Qy 1620 AATGCAACTGACCTCTCAGATAGGAGAGAGCGCTTCTCTTCTTCTACAGACCAAGTTTGT 1679
 Db 151463 -----GACTACTCAGACATAGGGGGTAACTTCTCAGTCTCTACAGACCAAGCTTGT 151413
 Qy 1680 AGCAACAAAATGATAGATATATTTTGGGAGGGGCCCCCAGTGACTTTGGAAGCCAGCTCT 1739
 Db 151412 ACCAACAAAATGATAGATATATTTTGGGAGGAGCCTCCAGTGACTTTGGACACCACTCC 151353
 Qy 1740 GATTTGAAAAGTGAGGATTTGGGCGGAGGAACTCTGAGGACGATGGCTTTGATAGCGATGGC 1799
 Db 151352 GATTTGAAAAGTGAGGATTTGGGATGAGGAACTCTGAGGATGATGGTTTGTAGTGTAGTGGC 151293
 Qy 1800 TCCTGTCTCAATCAGACGTTGGAACAGGACTTCGGAAGGCTTCCACTTTTGGAACTCTTTTC 1859
 Db 151292 TCCCTATCTGAATCAGACAGGAAACAGACTCCGAAGGGCTTCCACTTTTGGAACTCTTTTC 151233
 Qy 1860 CACAGTGTAGATCTTTACAAACCCCAAAATTTTACAGCCACGATTTACAGCGGCTGCCAGA 1919
 Db 151232 TACAGTGTAGATCTTTACACCCCCAGAACTTTTACAGCCACAATTCAGACTGCTGCCAGA 151173
 Qy 1920 ATTGCCCCCAGAGACCCATCAGATTCAGGACATTCCTGTGTCTGCGAGCTGTGTGTAGGG 1979
 Db 151172 ATTGCCCCCAGAGACCCATCTGATTCAGAGAGTCTCTGTGTCTGCAACTCTGATGTAGGG 151113
 Qy 1980 AGCTGTCCAGAGGAGCCCTTCCGAGAGACCCCAACCATAGTTTCCGGGAGGAAAGATGAC 2039
 Db 151112 AGTTCTCAGGCCACATCTCTTCCAGACCCCTTGACACAGATTTCCGGGAGGAGGATGAC 151053
 Qy 2040 TGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGGAATCTCTTCTGTCAATCT 2099
 Db 151052 TGGGAATCCAGTGCAGATGAAGCAGAAAACTTTAGATTTGTGGAATCTCTTCTGTAAATCT 150993
 Qy 2100 GAGAACCCCTACAACTTTTAAATTTTAAAGCTTCTTTTCAAAGCTCAGGGAAGAAATGG 2159
 Db 150992 GAGAACCCCTACAACTTTTAAATTTTAAAGCTTCTTTTCAAAGCTCAGGGAAGAGCTGG 150933
 Qy 2160 AAAGGCGTCCAGGACTCAAAAGGCTCTTCTGAGGTCAAGTGGCTTCTCTGGGCATCAT 2219
 Db 150932 AAAGGCGATC-----AGGCATCTCTGAGGCCACAGTGTCTTTTCTGGGCATCAG 150882
 Qy 2220 ACCTTACTTTTGTGAAGGCCAGCTGTGTAGAGGCGCAAGAGATAATTTGTCCAGGCTGT 2279
 Db 150881 ACCTTACTTTTGTGAGGCACAACTGTGTAGAGGCGCAAGAGCAATTTGTCCAGGCTGT 150822
 Qy 2280 GGGCTGGGTGAGGCTCTTGTCTGAGAAAGATACACCATATCAAGAGAAAAAGGTAAAC 2339
 Db 150821 GGGCTGGGTGAGGCTCTTCTGAGAAAGATACACCATATCAAGAGAAAAAGGTAAAGTGT 150762
 Qy 2340 TTCTTGGGAAGAGTTACTGAGTATATATAAGTGGTGTGATGAGATCCCAAGGA 2393
 Db 150761 TTCCATTTACATGACTATATATTTGTTGATATTTGATATGTTTACAGTAGAGA 150708

RESULT 8
 AC105827
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-101J7, *** SEQUENCING IN PROGRESS ***.

AC105827 248458 bp DNA linear HTG 21-SEP-2002
 Rattus norvegicus clone CH230-101J7, *** SEQUENCING IN PROGRESS ***.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Aclab (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNVL
Center clone name: CH230-101J7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 231489 bases at least Q40
Consensus quality: 232351 bases at least Q30
Consensus quality: 232940 bases at least Q20
Estimated insert size: 249328; sum-of-ctigs estimation
Quality coverage: 5x in Q20 bases; sum-of-ctigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the sizes
* is believed to be correct as given, however the pieces
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 248458: contig of 248458 bp in length.

Location/Qualifiers
1..248458
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-101J7"
6..560
/clone="clone_boundary"
clone_end:T7
site:EcoRI
end_sequence:BH312302"
complement(247315..248169)
/clone="clone_boundary"
clone_end:Sp6
site:EcoRI
end_sequence:BH312303"

misc_feature
misc_feature

ORIGIN

Query Match 30.9%; Score 1689.2; DB 2; Length 248458;
Best Local Similarity 84.5%; Pred. No. 0;
Matches 2022; Conservative 0; Mismatches 323; Indels 49; Gaps 9;
QY 1 CGGTCTCCGCTCGCCTGCGAGCTTCGGGTGCGGCTTCGGGCTTTGGCGCTTTTCGGGCTTCG 60
3946 CGGTCTCGTTCGTTCGTTCGAGCTTCGGGTTCGGGCTTCGGGCTTCGGGCTTCG 4005
61 TTCTTTTGGCGCTCGCTTCGCCACCCAGCCACCTTCCTTCGGCTTTGGCGCTTTTCGGGCTTCG 120
4006 TTCCTTTTCGCTTCGCCCGGCCACCCAG---CCCTAACGCTGTGTATTTTCGGGCTTCG 4061
121 TCGGGGCCACCGGAAACGCCGCCGCTGTCTCCGTCGGCGCGCGCGGAGGAGGCTTCT 180
4062 TGTGGGCGCG-----GGTCTCCGGAAACGTTGCCCGCGCGCGCTGTGAGGGGTCTTCT 4115
181 CTATGGTGGAGGATCTTCACAGCGCTTAGGACGCTCTCTTCCTTAGCGGGATGGACCTA 240
4116 CTATGGTGGAGGCTCTCACAGGACTAGGGCGCTTCAT--TTCGCGGAGATGGACGTA 4174

AC105827.3 GI:23265261
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 248458)
1 Mzrny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Gehr,K., Gill,R., Grady,M., Guerra,T., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kraft,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajda,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 248458)
Worley,K.C.
Direct Submission
Submitted (10-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 248458)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced q1:21736987.

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Qy	241	ACCGGCTGCGCAACCGCTTGCGCGGCGCTCTGGGCGGTCGGTGGAGCAGCATCTGTTGCGGA	300
Db	4175		
		GTCCGGTTCGCGACCGCTTGCGCGAGCCGCGCGGTACGTGTACATCGCTCGCTACGGA	4234
Qy	301	AGCCGCGCTCTCTGGGCGCTCTCTGCGGCGCGGGATCGGACTGCGAGTACCCACTCCG	360
Db	4235		
		AGCCGCGCTCTCTGGG---CCTCTGCTGGCGCGGGATCGGAATGCGATACCCACTCCG	4291
Qy	361	TGGCTGGCAAGCGGAGACTGTGTAGACCTCCGATCCAGCGCTCGGCTGACCGCGCTGAG	420
Db	4292		
		TGTCTGGCAAGGCTAGACTGTGTAGGCTCTCGATCTGCGCCGCTCTGACCGCGCTGAG	4351
Qy	421	CTCTGTCTCTCTCTGTCTGAGAACCGCCAAAGAAAGGAGATGGAGACGGAACGACACA	480
Db	4352		
		CTC---CGTACTCTGTGCCGAGAAGTTGCCAAGGAAGGATGGAGACGGGAACACACA	4408
Qy	481	GGGCGCGAAGCGGCTGGCCCTCGGCTGGGCTCCTGGTTCGGGCTGCCCTTCTTCCGCG	540
Db	4409		
		GGGCTCGGAAGCGGCTACCTCTCTCGGATGGGCTCTCTGGTTCGGGCTGCCCTTCCCTCGCG	4468
Qy	541	GATCGCAGCGCTCTCTTCGGAGTTCCGCGCGCTTCTCTCTCGACAAAAATCCCGGGAAC	600
Db	4469		
		GTTTCGACGCGCTGACTTCCGAGTTCGGAAGCCTTCTTCTCGACAAATCTCGAGTCT	4528
Qy	601	CCGCTCTGCCGAGCGTCGGAACGAGTACTGAGCAAAATGCTTCTCAGCTCCTTCGCC	660
Db	4529		
		CCGCCGCGCGAGCCTCGGACCAAGTACTGGACCAAAATGCTTTCAGCTCCTTCGCC	4588
Qy	661	TGCTCCCTAGCCTATTCCAGAAGCTGTGCTTTGGAGCGCAGCTTTCGGGGGCTGATTC	720
Db	4589		
		TGCTCCCTAGCTTGCTCCAGAAGCTGCTTATTTGGAGTCAGCTTTCGTTGGCATGATTC	4648
Qy	721	CTACCAGATGCTAGATTTTSCCGCAAGTTACGCGCTTGAGAGCTTCCAGAGGACGGG	780
Db	4649		
		CGACCAAGTGTAGATTTTTCAGCAGTTACAGCGCTCTAAGAGCTTTGAGAGGACGGG	4708
Qy	781	AGGAATCTGACGCTCCACGCTGAGAGTCTCTGAGTTACATCGCGCTGGACTCTTTCG	840
Db	4709		
		AGGAATCAGCGCTCCACAGTGCAGAAAATCTTTGAATT-CTCTCGGCTGGACTCTTCT	4767
Qy	841	CGAAGACTCGGCGTCTCAGTACTCTTGCAATGGCTAGAGAGGACTTCAGTGGCAGTGC	900
Db	4768		
		TCGGAAGACTTGGTTGTTCAGTTCCTTTGATTTGGCTAGAGGAGGACTCCAGGGGACGTGC	4827
Qy	901	TCGTCCTCAGA-CTGGAGTTAAACTCAAGGCCACAGGAAGAGCTTTAGACTCTGACGC	959
Db	4828		
		TCGTCTCTCAGACCTGGAAATGAAATCAAGGCCACAGGAAGAGCTTTAGACTCTACAGC	4887
Qy	960	CCCACTTTCCTCTCGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCTTCAA	1019
Db	4888		
		CACACTTTCCTCTTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCGGTAGTCTTCAA	4947
Qy	1020	GCTGGTCTAGTCTCCACCGAGAACTTGAATCTTCATCCTCTGGGCTCTGAGGGTTTCAG	1079
Db	4948		
		ACCGETTAGTCTCCACCGAGAACTGGCTCTTCATCCTCTGGGCTCTGAGTACTCAG	5007
Qy	1080	AGCTTAGTAAATTTCAAGTAGTTTCTATCTCTGGAACCTTCTACTCTGGACTACTT	1139
Db	5008		
		AGCTTGGGCAATTTCAAAGTAGTTTCTATCTCTGAAACCTTCCCTACTGGAATGCTT	5067
Qy	1140	CCCCAGTTTAGGCTCGCGCTGTACAGACAGCGCTCGAGGTGGCAAGTTGTGGGTTTCCGA	1199
Db	5068		
		CCCCAGTTAGAGCTGCGCTGTCAAGACAGGTGGAGGTGGCCAACTTGTGGATTTCCGA	5127
Qy	1200	ACACTAAACCCAGAGAGTGCTATCTTTCTGAAGATGGTTGTCACTCTCAGCCGTTGCGG	1259
Db	5128		
		ACACTATCCCAAGAGCTGACTCTTCTGAAAGTGTGTCACTACCTCAGCCGCTGTGT	5187
Qy	1260	GCAGAGATGCGGCAACCGCTCGAGAGGTGTCGCGCTCTCTCTACAGAAAGCTGCGG	1319
Db	5188		
		GCTCAGATGTCGGCAATGCGCTGCGGAGGTGTCGCGCCCTCTCTACAGAAAGGCTGCCA	5247
Qy	1320	GAATCCACACCTTCGTATGAAACGGCTAGAAATTCCTCCAGGCTAAACAAAGGCAAGAG	1379

BD160691
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BD160691
Primer for synthesizing full-length cDNA and use thereof.
BD160691
BD160691.1 GI:278666449
JP 2002191363-A/15534.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 2942)
Oka, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, I.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 15534 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/15534
PD 09-JUL-2000
PF 28-JUL-2002 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS
Location/Qualifiers
1..2942
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source

ORIGIN

Query Match 24.2%; Score 1320.8; DB 6; Length 2942;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 162; Conservative 0; Mismatches 662; Indels 84; Gaps 16;

Qy 151 CGGTCGGCGCGCGAGGAGGTCTTCTCTATGGTGGAGCGATCTCACACGGCCTTAGG 210
Db |||||
96 CCATCGCGAGGCGGTGGGGGACTCTATCCATGGTGTGAAGCGCTGAGCCGACTAGG 155
|||
Qy 211 ACGTCTCTTCCCTAGCGGGATGGACCTAACCGCGGTGCGCACCGCTTGGCGGGCGCTC 270
Db |||||
156 GAACCTCTCTTCCC -CGGCAGAGATGGAGTCGATCAGTCGCGCGCTATTGCGCGGGTGT 214
|||
Qy 271 TGGGCGGTGCGGTGACGACCTCGTTGCGGAAGCGCGCTCTCTGGGCTCTCTGCGGG 330
Db |||||
215 TCCTTCCCTGTGTTCTGGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTCTGCTGG 274
|||
Qy 331 CGCGGGAATCGAGCTGACGTACCCACTCCGTGGCTGGGCAAGCGGAGACTGTGTAGACC 390
Db |||||
275 CTCGAAGATCGCCTGGAGCAGACGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334
|||
Qy 391 TCGGATCCAGCTCGGCTGACGCCGCTGAGCTCTGTCTCTCTCTCTGTGTGAGAACGCCGC 450
Db |||||
335 TCCTCCGAATCCCGTCCGACCTCCAGCGCTGAGCGCGCGCGCGCTACCTGAGAGACTGTC 394
|||
Qy 451 AAG-GAAGAGATGGAGACAGGAACGCCACAGGGCCCGGAAGCGCCTTGGCCCTCGGCTG 509
Db |||||
395 AAGAAAAGGAGATGGAGCGGGGACAGCGCGATCGCGGAACCGCTTGGCCCTCGGGCG 454
|||
Qy 510 GGCTCTCTGGTTCCGGCTGCCC - - -TTCTTCGGGATCGCACGGCTGTCTTTCGGAGTTC 566
Db |||||
455 GGCTTCGGTTCTGGCCACCCCTTTTCCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTC 514
|||
Qy 567 CCGCGCGCTTCTCTCGACAAAATCCCGGGAATCC - - - - -GCTCTGCCC 611
Db |||||
515 CCGACGCTCTTGGCCCGGGAATCTCCGGGAACCCACACATGCTTCTCTGCGCCAGGCC 574
|||

QY 1661 TTCTACCAAGCAGTTGTAGCAACAATGATAGATTATATTTTGGAGGCGGCCCCOAG 1720
Db 1651 TTCTCCAGACCAGCTTGTAGTAAACAATCTGATAGATTATATTTTGGAGGTGATCCAG 1710
QY 1721 TGACTTGGAGCCAGCTCTGATTCTGAAGTCAGGATTGGGCGGAGGAACCTGAGGACGA 1780
Db 1711 TGACTTGGAAACAAGTTCTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATGA 1770
QY 1781 TGGCTTTGATAGCAGTGGCTCCTCTCTGAATCAGACCTGGAACAGGACTCGGAAGGCCT 1840
Db 1771 TGGTTTGTAGTAGTAGTCACTGTCTGAGCTCAGACCTTGAACAGACCTCGAAGGCCT 1830
QY 1841 TCACCTTTGGAACTCTTTTCCAGAGTGTAGATCTTTTACAAACCCCAAACTTTTACAGCCAC 1900
Db 1831 TCACCTTTGGAACTCTTTCTGAGTGTAGATCTTTTATAATCCCAAGAACTTTTACAGCAAC 1890
QY 1901 GATTACAGCGCTGCCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGAGCATCTGTGTC 1960
Db 1891 AATTCCAGACTGCTGCCAGAAATTTGTTCTTGAAGAGCCTTCTGATTCAGAGAAAGGATTTGTC 1950
QY 1961 TGGCAGCTGTGTGTAGGGA---GCTGTCCAGAGGACCCCTTCCGAGAGACCCCGACCA 2017
Db 1951 TGGCAAGTCTGATCTAGAGAAATCTCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCA 2010
QY 2018 TAGTTCCGGGAGGAAGATGACTGGGAACCCAGTGCAGATGAAGCAGAGAAATCTTAAAT 2077
Db 2011 TAGTTCTGGGAGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGATCTCAAACT 2070
QY 2078 GTGGAACCTTTCTGTCTGATCTGAGGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTT 2137
Db 2071 GTGGAACCTCATTTCTGTAAATCTGATGACCCCTTACAACTTTTAAATTTTAAAGGCTCCTTT 2130
QY 2138 TCAACCGTCCAGGAAGAAATTTGGAAGGCGCTCAGGACTCAAGGCTCTTCTGAGGTCCAC 2197
Db 2131 TCAAAATCATCGGGAAATGAGAAAGGCTGTGTGACTCAAAAGCCCATCTGAGTCCAT 2190
QY 2198 AGTGCCCTTCTCTGCCCATCATACCTTCTTGTGAAGCCCAAGCTGTTTAGAGAGCA 2257
Db 2191 TGTGGCCATTTCTGAGTGTCAACCTTACTTTCTTGTGAAGTGCAGCTGTTGGGGAGCA 2250
QY 2258 AGAAGATAATTTGTCAGGCTGTGGCTGGGTGAGGCTTCTCTGGAGAAAGATACACCCA 2317
Db 2251 AGAAAGTGAATTTCCAGACTCGGTACAGCGTGAAGTCTTCTTGGAGAAAGACACACA 2310
QY 2318 TATCAAGAGAAAAAGGTAACTTCTCTGGAAGAGTACTGAGTATTATATAAGTGGTGA 2377
Db 2311 TGTCAAAAGAAAAAGGTAACTTCTTCTTGAAGAGTACTGAGTATTATATAAGTGGTGA 2370
QY 2378 TGAGGATCCGAAGGACCATGGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2437
Db 2371 TGAGGATCCGAAGGACCATGGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2430
QY 2438 AATTCAGAAACAGAAAGTTGCGATTTGGCTACTGCTTGGCCCTTTGAGCACAGAAAAAT 2497
Db 2431 AATTCAGAAACAGAAAGTGTATTGGATATTGCTTGAACATTTGAACACAGAGAAAGAT 2490
QY 2498 GTTTAATAGACT---GAGGATCGAGTCAAAAGGACTTACTGTGTGTACAGCAATGTTAAGA 2553
Db 2491 GTTTAATAGACTCCAGGGAACATGCTTCAAAAGGACTTAATGTCTCAAGCAATGTTAGT 2550
QY 2554 AGTGA-ACAGCCTGCAACCCGTCGCCACTCTGTCTCTTACTTGTAGA-GTTTCCCTTAAA 2611
Db 2551 TGGCAGCCTGTAGTCTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTTTTAAAA 2610
QY 2612 ACAAACTGGCAGCTGCTTGGACATGTTTTTAAAGAAACAACCTGTATCTAGAGATG 2671
Db 2611 ACAAACTTGGCAGCTGCTTGGACATGTTTTTAAAGAAACAATTTTATCTTGGATCT 2670
QY 2672 CAGTTTGAATTTTGGGTAAATGCTCATTTAGAA-----ACACCAACT 2717
Db 2671 AGTTTAAATTTTTTTTTTGGAAACATATCCCACTCAGAAACAATTCAGGTTTGAAGCCGCC 2730
QY 2718 CCGATAATGAAGAAT-----CTCTTATCTGTAAATCTCTCTCTTTTCTCTTATTTAGTGGAT 2771

Db 2731 CTGATAATGAAGATGAAGTAGTGTGATTTCTAATCTCCTCTTTTGTATTAGTGGAT 2790
QY 2772 GTGGGTTT 2779
Db 2791 GTGCTTTT 2798
RESULT 10
AX884059
LOCUS AX884059 2942 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 18964 from Patent EP1074617.
ACCESSION AX884059
VERSION AX884059.1 GI:40038960
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Sugai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18964 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2942
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
407..2548
/note="unnamed protein product"
/codon_start=1
/protein_id="CA892217.1"
/db_xref="GI:40038961"
/translation="MEPTGSGSRKRLGPRAGFRFWPFFRRRSQAQSSKEPTPLGPEN
SGNPTLLSSAQPETRVSYMTKLISQLLAPLGLQLKVLWSLPGMFPFRWLDFAGV
YSALRALKREKPAFTAQKSLSSQLDSDSPVSTPLDWLEIGHWSPYPPDLKLEL
KAGSALDPAQAQLLEQLWGVYLLPSLSQRLYSNRELSSGSPGFLNQRIDDFSV
VSYLLNPVLDLCPFLVSYNSGNSVVGFTLTPSSCLREDHCHGPOLGAEILP
ASWQGPCPLSTEGLEPIHLRMKRLFLQOASKQDLPTPDQNGVHSLREHSLLRM
DPKXDRNDPQFVPAAGDIPGNTQESTERKIELLTTEVELALEEESPSCPSSEIPM
EKSPGGRISVDVSYLEGLDPLISARPACSNKLIDYILGASDSLETSSDPGEWDME
EABDDGFDSDSLSDLEQDPEGLHWNFSFCLVDPYNQNFATTTQTAAARIVPDEES
DSKFLSGKSDNSQSGSLPETPSSGSEDDWESSADEASFKLWNSFCNDSQVQ
NPLNFKAPFOTSGENKGRDSTPSESIVAISECHTLLSKVQLLGSQSEPCDSYQ
RDVLSGRHTHVKKVTFLEVTYVYISGDEDRKGFWEFARDGCRFQRIQETEDA
IGCLTFEHRHRMFRNLQGTCTCFGLNVLKQC"

Query Match 24.2%; Score 1320.8; DB 6; Length 2942;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 15;
QY 151 CCCTGCGCCGCGCGAGGAGGCTCTTCTCTATGTTGGAGCGATCTCACACGCGCTTAGG 210
Db 96 CCATGCGCGAGGCGGTGGGGGAGCTCTATCCATGTTGTTGAAGCGTCGAGCGACTAGG 155
QY 211 ACCTCTCTTCCCTAGCGGGATGAGACTAACCGCGTCCGACCGCTTGGCGGGCCTC 270
Db 156 GAACCTCTTCCC-CGCGAGGATGGAAGTCGATCAGTCGCGCTTATTCGCGGGCTGT 214
QY 271 TGGGCGCTCGCGTGCAGACTCGTTGCGGAAGCGCGCTCTCTGGGCGCTCTCTGCGCG 330
Db 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGG 274
QY 331 CGCGGGAATCGAGCTGAGTACCACTCTCGTGGCTGGGCAAGCGGAGAGCTGTGTAGACC 390
Db 275 CTCGAGATCGGCTGAGCAGCGACGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334
QY 391 TCGGATCCAGCTCGGCTGAGCGCGCTGAGCTCTGTCTCTCTCTCTGCTCAGAGAGCGCC 450
Db 335 TCCTCCGAATCCCGCTGACCTCTCAGCGCGCTGAGCGCGCGCTTACCTGAGAGACTGTC 394

Db 2551 TGGCAGCCTAGTCTAGCTAGCATACACATCTCTTACCTGAGAGGTGCTTTTAA 2610
Qy 2612 ACAACACTGGCAGCTGCTTGGACATGTTTTTAAAGAAACAACCTCTACTAGAGATG 2671
Db 2611 ACAATCTTGGCAGCTGCTTGGACATGTTTTTAAAGAAACAACCTCTACTAGAGATG 2670
Qy 2672 CAGTTTGATTTATTTTGGGTAAATGTCTCTATTAGAA-----ACACCAACT 2717
Db 2671 AGTTTAATTTTATTTTGGGTAAATGTCTCTATTAGAA-----ACACCAACT 2730
Qy 2718 CGATAATGAAGAT-----CTCTATCTGTATTCCTCTCTTTTCCTATTAGTTGAT 2771
Db 2731 CTGATAATGAAGATGAAGTGTGTGATTTCTTAATCTCTCCCTTTTTCGATTTAGTTGAT 2790
Qy 2772 GTGGGTTT 2779
Db 2791 GTGCTTTT 2798

RESULT 11
AK027650
LOCUS Homo sapiens cDNA FLJ14744 fis, clone NT2RP3002770. 2942 bp mRNA linear PRI 30-JAN-2004
DEFINITION AK027650
VERSION AK027650.1 GI:14042484
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, Y., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosiiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yanazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotta, Y., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Toqiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y.,
Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Negase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)

2
14702039
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,
Tanase, T., Nomura, Y., Toqiya, S., Komai, F., Hara, R., Takeuchi, K.,
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.
NEO human cDNA sequencing project
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
source

1. .2942
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3002770"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/notes="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
407. .2548
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB55266.1"
/db_xref="GI:14042485"
/translation="MBPFGTGGSRKRLGPRAGFRFWPFFPRRQAQSSKPTPLGPEN
SGNPTLSSAQPTETVSYNTKLLSLLQLLPGLLQKLIWSQLFGMFPTRWLDFAV
YSALRALKGRKPAAPTAQKSLSSLDSDSPSVTSPDLWEIGIHQWYSPDLKLEL
KAGSALDPAQAFLLEQOLGWVLLPSSLSRLYNSRELSSPSPLNQRIDDFSV
VSYLLNPSYLDLCPRLVSVQNSDGNSEVVGFTLTPESSCLREDHCHPCLARLIP
AQWGGPPLSTEGPIHHLMKRLEFLQASXGQDLPTPDQNGHSLSEESLLAM
DPKHNDPTQFVPAAGDIPGNTETEEKIELLTTEVPALAEESPSBEGCPSSEIPM
EAEEDGDSLSLDLEQDPEGLLWNSFCVSPYNPQNTATQTAAARIVPEEPS
DSKDLGKSLDENSESQSLPTPEHSSGEDEDDWESSADEASLKLWNSFCNDDPY
NPLFKAPQVTGSENGKCRDSTPSESVIAISECHTLKSVQLLGSQSSCEPDSVQ
RDVLSGRTHVRKRVFLVEVTEYIISGDEDRKGPWEFEFARDGCRFQRRIQETEDA
IGYLTFFEHRRMFNRLQGTGCFKGLNVLKQC"

CDS

ORIGIN

Query Match 24.2% Score 1320.8; DB 9; Length 2942;
Best Local Similarity 72.5% Pred. No. 0;
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;
Qy 151 CCCTGCGCCGCGCGGAGGAGGTCTTCTTATGTTGGAGCGATCTCACACGCGCTAGG 210
Db 96 CCATGCGCGAGGCGGTGGGGGAGCTCTATCTCATCTGTTGAAGCGTGGAGCGCTAGG 155
Qy 211 ACCTCTCTTCCCTAGCGGGATGGACCTAACCGCGTCCGCCACCGCTTGGCGGGGCTC 270
Db 156 GAACCTCTCTTCCC-CGCGAGGATGGAAGTGCATCATGTCGCCGCTTATTCGCGGGGTGT 214
Qy 271 TGGGCGCTCGCGTGCAGCACTCGTTGGGAAGCGCGCTCTCTGGGCGCTCTCTCGCGG 330
Db 215 TCTTCTCTGTGTTCTGCGCGCGCGCTGCGGATTTGCTGCGCTCTGTGGCTTTCTGTGG 274
Qy 331 CGCGGGAATCGGATCGAGTACCCTTCCGTGGTGGGCAAGCGGAGAGCTGTGTAGACC 390
Db 275 CTCGAGATCGGCTCGGAGCAGCGACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334
Qy 391 TCGATTCAGCTCGCTGAGCGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAGCGGCC 450
Db 335 TCCTCCGAATTCCTCGCTCGACTCTCAGCGCGCTGAGCGCGCGGCTTACCTGAGAGCTGC 394
Qy 451 AAG-GAAAGGAGATGGAGACAGGAAACGACAGCGCGCGGAGCGGCTTGGCCCTCGGCTG 509
Db 395 AAGAAAAGGAGATGGAGCGCGGAGCAGCGGATCGCGGAAACGGCTTGGCCCTCGGCGG 454

Qy 510 GGCTCTCGTTCGGCTGCCCC---TTCTTTCGGGATCGCAGCGCTGCTCTTCGAGTTC 566
Db 455 GGCCTTCGGTTCGGCCACCTTTTTCCTTCGGGATCGCAAGCAGGCTCTCTTAAGTTC 514
Qy 567 CCGCGCTTCCTCTCGACAAATCCCGGAATCC-----GCTCTGCC 611
Db 515 CCAGCGCTCTTGGCCCGGAATCTCGGAAACCCACACTGCTTTCCTCTGCCAGCCC 574
Qy 612 GAGCGTCGACCGAGTACTCGAACAAATTTCTCAGCTCCTTGCCTGCTCCCTAGC 671
Db 575 GAGACTCGGTCAGTTACTGACGAACTGCTCTCCAGCTCCTTGGCCGCTCCCGGA 634
Qy 672 CTATTCAGAGCTGCTGCTTTTGAGCCAGCTTTCGGGGGCTGATCTCTACAGATGG 731
Db 635 TTGCTTCAGAGGTGCTAAATTTGAGCAAATTTTTCGGTGGAAATGTTTTCGACCAATGG 694
Qy 732 CTAGATTTTCCCGCAAGTTACAGCCCTCAGAGCTTCGAGAGCAGGGAGGATCTGAC 791
Db 695 CTAGATTTTCTGAGTCTACAGCGCTCAGAGCTCAGAGGACCGGAGAAACCGACC 754
Qy 792 GCTCCACCGTGCAGAACTCTCGAGTTACACTCGGCTCGACTCTTCGCGAAGACTCGC 851
Db 755 GCCCCACAGCGCAGAAATCTTTGAGTT-CGCTGCACTCGACTCCTC---AGACCCCTC 810
Qy 852 GTCGTCACTACTCTTGCAATGGCTAGAGAGGACTCCAGTGGCAGTGTCTGCTCTCAGA 911
Db 811 GGTCAACAGTCCCCTTGAATGGCTAGAGAGGAGTCCACTGGCAATCTCGCCCCCAGA 870
Qy 912 -CTGGAAGTTAACTCAAGCCCAAGGAAGCTTTAGACTCTCGACGCGCCACTTTCCT 970
Db 871 CCTAAATTTGAGCTTAAGGCCAAGGGAAGTCTTTGGACCTCGCAGCAAGGCTTTTCT 930
Qy 971 CTTGGAGCAGCAGCTGTGGGGAGTGGAGTTGCTGCCCAGTAGCCCTCAAGCTGCTTAGT 1030
Db 931 CTTAGACAGCAGCTGTGGGGAGTGGAGCTGTTGCCCAGTAGCTTCAATCCCGTCTGTA 990
Qy 1031 CTCCACCGAGAACTTGACTCTTCATCTCTGGGCCCTCTGAGCGTTGAGAGCTTAGTAA 1090
Db 991 CTCTAACCGGAACTTGGCTCTTCGCCCTCTGGGCCCTCTAAACATTCAGGCACTAGACGA 1050
Qy 1091 TTTCAAGTATGTTCTCTCTGACCTTCTCTGACCTTCTCTACCTGCACTACCTCCCGAGTTAGG 1150
Db 1051 TTTCAAGTATGTTCTCTCTGACCTTCTCTGACCTTCTCTACCTGCACTACCTCCCGAGTTAG 1110
Qy 1151 GCTGCGCTGTGAGAGCAGCTGTGAGGTGGCCAGTTTGTGGGTTTCCGAACTAAACCCC 1210
Db 1111 AGTCAGCTATCAGAACAGTATGAAATAGGAGGTAGTGGCTTCCAGACACTAACCCC 1170
Qy 1211 AGAGAGCTGTATCTTTCTGAAGATGTTGTCACTCAGCCCTCAGCGTTGCGGCGAGATGTC 1270
Db 1171 AGAGAGCAGCTGCTGAGAGAGGACCAATTTGTATCCCGCAGCGCTGAGTGCAGAACTCAT 1230
Qy 1271 GGCNACCGCTGAGAGAGTGTGCGCTCTCTCTACAGAGGCTGCGGAAATCCACCA 1330
Db 1231 TCCGCGCTCTGCGGAGGATGTCCACCTCTTCTACGGAAGGCTTACAGAAATTCACCA 1290
Qy 1331 CCTTCGTATGAAACCGCTAGAAATTCCT---CCAGGCTAAACAAAGGCAAGAGTTTACCCAC 1387
Db 1291 TCTTCGATGAAACCGCTGGAATTCCTTCAAGGCTAGCAAGGGCGAAGATTTACCCAC 1350
Qy 1388 CCCTGACCAAGATATGCTATCATAGCTGGAGAGGAAATAACTTCTTCGGATGGA 1447
Db 1351 CCCTGACCAAGATATGCTATCAAGCTGAGAGGAAACACAGCTTCTCCGGATGGA 1410
Qy 1448 CCCAACATTTGACAGATTAACCCAGCAGCGCGTGTCCCTGCTGCGACAGGCT--- 1504
Db 1411 TCCAAACACTGACAGATTAACCCAAACAGATTTGTTCTCTGCTGCTGGACATTTCTCTGG 1470
Qy 1505 -----GGAGCCCACTGAGAAACCAAGAAATTTGATTCAGAAAGTTTTCAC----- 1551
Db 1471 AAACACCCAGGATTCCTGAGAGAAATTAAGAAATTAATTAACACAGAGTTTCCACTGCG 1530
Qy 1552 -----AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAATTAACCGTGGGAAAGAA 1600

Db 1531 TTTGGAAGAGAGAGCCCTTCTGAGGCTGTTCATCTAGTGAGTACCTATATGAAAAAGGA 1590
Qy 1601 ATGTGAAGAGGACCAACATAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCCTTCTGT 1660
Db 1591 GCCTGAGAGGCGGAATAAGTGTAGTTACTCATACCTAGAAAGGTGACCTTCCCAT 1650
Qy 1661 TTCTACAGACCAAGTTTGTAGCAACAACTGATAGATTAATATTTTGGGAGGCGCCCCAG 1720
Db 1651 TTCTGCGACAGCAGCTTGTAGTAACAACTGATAGATTAATATTTTGGGAGGTGCATCCAG 1710
Qy 1721 TGACTTTGGAAGCCAGCTCTGATTTCTGAAAGTGAAGGATGCGGCGAGGAACTCGAGACGA 1780
Db 1711 TGACCTTGGAAACAAGTTCTGATCCAGAGGTGAGGATGCGATGAGAAAGCTGAGATGA 1770
Qy 1781 TGGCTTTGATAGCGATGGCTCCCTGTCTGAAATCAGACGCTGGAACAGGACTCGGAAGCCT 1840
Db 1771 TGGTTTGTATGTATGTACTCTGTCTGAGCTCAGACTGAGACCTTGAACAAGACCTCGAAGGCT 1830
Qy 1841 TCACCTTTGGAATCTTTTCCAGTGTAGATTCCTTACAAACCCCAAACTTTTACAGCCAC 1900
Db 1831 TCACCTTTGGAATCTTTTCTGAGTGTAGATTCCTTATAATCCCCAGAACTTTTACAGAAC 1890
Qy 1901 GATTGACGCGCTGCCAGAAATGCCCCCAGAGACCCATCAGATTCAGGAGACATCTCTGGTC 1960
Db 1891 AATTGACAGCTGCTGCCAGAAATGTTCTGAAAGAGCCTTCTGATTCAGAGAAGGATTTGTC 1950
Qy 1961 TGGCAGCTGTGGTGTAGGGA---GCTGTGAGGAGGAGCCCTTCCCGAGACCCCCGACCA 2017
Db 1951 TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCA 2010
Qy 2018 TAGTTCGGGGAGGAAAGATGACTGGGAAACGAGTGCAGATGAAAGCAGAGAACTCTTAAAT 2077
Db 2011 TAGTTCGGGGAGGAAAGATGACTGGGAAATCTAGTGCAGATGAAAGCAGAGAGTCTCAAACT 2070
Qy 2078 GTGGAATCTTTCTGTCTCTGAGGACCCCTACAACTTTTAAATTTTAAAGCTCCTTT 2137
Db 2071 GTGGAATCATTTCTGTAAATCTGTATGACCCCTTAAACCTTTTAAATTTTAAAGCTCCTTT 2130
Qy 2138 TCAACCGTTCAGGGAAGAAATGGAAAGCGCTCAGGACTCAAGGCTCTTCTGAGGTGAC 2197
Db 2131 TCAAACATCAGGGGAAATGAGAAAGGCTGTCTGACTCAAGACCCCATCTGAGTCCAT 2190
Qy 2198 AGTGGCTCTCTCTGGCCATCATACCTTACTTTCTGTAAGGCCAGCTGTGTAGAGACCA 2257
Db 2191 TGTGGCCATTTCTGAGTGTACACCTTACTTTCTTGAAGGTGAGCTGTGTGGGAGCCA 2250
Qy 2258 AGAAGATTAATTTCTCAGGCTGTGGGCTGGGTGAGCTCTGCTGGAGAAAGATACACCCA 2317
Db 2251 AGAAAGTGAATGTCCAGACTCGGTAAGGCTGACGTTCTTCTGGAGGAAAGACACACA 2310
Qy 2318 TATCAAGAGAAAAAGGTAACCTTCTGGAAGAAAGTTACTGTAGTATATATAAGTGTGA 2377
Db 2311 TGTCAAAAGAAAAAGGTAACCTTCTTGAAGAAAGTTACTGTAGTATATATAAGTGTGA 2370
Qy 2378 TGAGGATCGCAAGGACCAATGGGAAAGAAATTTGCAAGGATGAGATGAGGTTCCAGAAACG 2437
Db 2371 TGAGGATCGCAAGGACCAATGGGAAAGAAATTTGCAAGGATGAGATGAGGTTCCAGAAACG 2430
Qy 2438 AATTCAAGAAACAGAAATTTGCCATTCGCTACTGCTTGGCTTTGAGCACAGAGAAAGAT 2497
Db 2431 AATTCAAGAAACAGAAATGCTATTTGGATATTCCTTGACATTTTGAACACAGAGAAAGAT 2490
Qy 2498 GTTTAATAGACT---GAGGATCGAGTCAAGAGGACTTACTGTGTGTACAGCAATGTTAAGA 2553
Db 2491 GTTTAATAGACTCCAGGAAACATGCTTCAAGGACTTAAATGTTCTCAAGCAATGTTGAGT 2550
Qy 2554 AGTGA-ACAGCTGCAACCGTGCCTGCTGTCTCTTACTTGAGA-GTTTCCCTTAAAA 2611
Db 2551 TGGCAGCTGTAGTCTCTAGCTAGTACATACCTCTTACCTGAGAGGTGCTCTTTAAAA 2610
Qy 2612 ACAAACACTGGCAGCTGCTCTTGGACATGTTTTTAAAGAAACACTTGTATCTAGAGATG 2671

Db 2611 ACAAATCTGGCAGCTGCTCTTGCATATTTTTTTTTTTAGAGAAATGTAATCTGGATCT 2670
QY 2672 CAGTTTGATATTTTTGGTAATGCTCATAGAA-----ACACCAACT 2717
Db 2671 AGTTTAAATTTTTTTTGGACATATCCCACTGAGAAACATTCAGGTTTGAGCCAGCC 2730
QY 2718 CGATAATGAGAAT-----CTCTTATCTGTAATCTCTCTTTTCCATTAGTTGAT 2771
Db 2731 CTGATAATGAGGATGAAGTGTGTAATCTCTTAATCTCTCCCTTTTTCATTAGTTGAT 2790
QY 2772 GTGGGTTTT 2779
Db 2791 GTGCTTTT 2798

RESULT 12
BC065280
LOCUS BC065280 5264 bp mRNA linear PRI 12-OCT-2004
DEFINITION Homo sapiens cDNA clone MGC:74824 IMAGE:6172811, complete cds.
ACCESSION BC065280
VERSION BC065280.1 GI:40807094
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 5264)
AUTHORS Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Dietzenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Tohiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huly, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

2 (bases 1 to 5264)
DIRECTOR MGC Project.
DIRECT SUBMISSION
TITLE Submitted (12-JAN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
JOURNAL NIH-MGC Project URL: <http://mgc.nci.nih.gov>
PUBMED Contact: MGC help desk
REFERENCE Email: cgapbs-remail.nih.gov
AUTHORS Tissue Procurement: ATCC
TITLE cDNA Library Preparation: Life Technologies, Inc.
JOURNAL cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC), National Institutes of Health Intramural Gaithersburg, Maryland.
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov

REMARK Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.I., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McBowell, J., Pearson, R., Stantripop, S., Thomab, P.J., Touchman, J.W.,

Tauregon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 140 Row: p Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 41282220.

FEATURES

source

1..5264
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:74824 IMAGE:6172811"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_67"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
381..2522
/codon_start=1
/product="Unknown (protein for MGC:74824)"
/protein_id="AAH65280.1"
/db_xref="GI:40807095"
/translation="MEFGTGGSRKRLGPRAGFRWPPFRRSQAGSSKPTPLGPEN
SGNPTLLSSAQPETRVSYWTKLLSLQLLAPLGLQLKVLWLSQFGMFPRLWLDFAV
YSALRALKGRKPAAPTAQKLSLSQLDSDSPSVTPDLWLEGIHWQYPPDLKLEL
KAKGALDPAQAFLLEQLWGLVGLPSSLSQSLRYNRELSSPSGSPQNLTRIDNFSV
VSVLQCPPLSTEGLEPIHLRWKRLBFLQOASKGDLPTPDQNGHSLREHSLLRM
ASWQGGPPLSTEGLEPIHLRWKRLBFLQOASKGDLPTPDQNGHSLREHSLLRM
DPKCRDNDPQFPAAGDIPQNTQSTEEKIELLTTEVFLALRESPSECPSEIPLM
EKEPGRISVDVSYLEGLDPLISARPACSNKLIDYLGGASDSDLETSPDEGEDWDE
EADDDGFDSDSLSDLEQDPGLHLWNSFCSDVPYNQNFATTTQTAARIVPEEPS
DSKDLKAPFOTSGENKGRDSETPSEIIVASECHTLKSVOLLGQSECEPDSVQ
RDVLSGRHTRHVKKVTLEEVTEYVIGSDEDEKGFWEFARDCGRFQKRIQETEDA
IGCLIFHRERNFNLQGTFCFGLNVLKQC"

CDS

ORIGIN

Query Match 24.2%; Score 1320.8; DB 9; Length 5264;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;
QY 151 CGGTGCGCGCGCGGAGGAGGCTCTTCTTATGTTGGAGCGCATCTCACACGGCTTAGG 210
Db 70 CCATCGCGGAGCGCGTGGGGGAGCTCTTATCCATGCTGTGTAAGCGCTCGAGCGGCTAGG 129
QY 211 ACGTCTCTTCCCTAGCGCGGATGGACCTAACCGCGTCCGCCACCGCTTCGGGGGGCTC 270
Db 130 GAACCTCTTCCCTCC - CGCAGGATGGAAGTCCGATCAGTCCGCCCTATTTCGGGGGCTGT 188
QY 271 TGGGCGCGTCCGCTGCAGCACTCGTTGCGGAAGCGCGGCTCTCTGGGCTCTCTTGCCTGG 330
Db 189 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCCCTCTGTGGCTTTCTGCTGG 248
QY 331 CGCGGGAATCGGACTGCGAGTACCACCTCGCTGGCTGGGGAAGGGGAGACTGTGTAGACC 390
Db 249 CTGGAAGATCGGCTTGAGCAGCGACGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 308
QY 391 TCGGATCCAGCTCGGCTGAGCGCGCTGAGCTCTGTCTCTCTCTCTGCTCAGGAAGCGCC 450
Db 309 TCTTCCGAATCCCGTGCACCTCCAGCGCGCTGAGCGCGCGGCGCCCTACCTGAGAGACTGC 368
QY 451 AAG-GAAAGGAGATGGAGACAGGAAACGCA CAGGCGCGCGGAAGCGGCTCGGCCCTCGGCTG 509
Db 369 AAGAAAAGGAGATGGAGCCGGGACAGCGGATCGCGGAAACGGCTTGGCCCTCGGGCG 428
QY 510 GGCTCTCTGTTCCGGCTGCCCC-----TTCCTTGGCGGATCGCACCGCTGCTCTTGGAGTTC 566
Db 429 GGCTTCGGGTTCTTGGCCACCTTTTTCCTCTCGGCGATCGCAAGCAGGCTCTTCTTAAGTTC 488
QY 567 CCGCCGCTCTCTCTGCACAAAATCCGGGAACTCC-----GCTCTGCTCCC 611
Db 489 CCGACGCTCTTGGCGCGGAAACTCGGGAAACCCACACTGCTTTCTCTGCGCCAGCCC 548

Db 2705 CTGATATGAGGATGACTAGTGTGATTTCTATCTCTCCCTTTTGTGATTAGTTGGAT 2764
 Qy 2772 GTGGGTTT 2779
 Db 2765 GTGCTTTT 2772

RESULT 13
 AL606489/c
 LOCUS Human DNA sequence from clone RP11-739N20 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL606489
 VERSION AL606489.26 GI:24940211
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 167079)
 REFERENCE
 AUTHORS Pearce,A.
 TITLE Direct Submission
 JOURNAL Submitted (08-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk
 On Nov 12, 2002 this sequence version replaced gi:22204251.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chri>
 RP11-739N20 is from the library RPCI-11.3 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

FEATURES
 source
 1. 167079
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-739N20"
 /clone_lib="RPCI-11.3"

ORIGIN
 Query Match 20.6%; Score 1124.8; DB 9; Length 167079;
 Best Local Similarity 71.4%; Pred. No. 9e-273;
 Matches 1686; Conservative 0; Mismatches 607; Indels 67; Gaps 13;

Qy 39 CTGCGGCATTTTGGAGCTTCGCTTCTTTTGGCGCCTCGCCTGCCACCCAGCCACCCCTTTCC 98
 Db 63783 CTGCGGCATTTTGGGCTTCGCTTC-----CACGCGACCGCGGCTTACCCAGTCTCTCC 63728
 Qy 99 GCCTTGGCGTTTCGCGCCTCCGTCGCGGCCACCGGAAACCGCGCGTGTCTCCGTGCGC 158
 Db 63727 GGTATCGCTGTCTCAGGGGCTTTTCAACCTCTGTCTCAGTCGG-----AAAACCATCGCC 63673
 Qy 159 CCGCGCGGAGGAGGGGTCTTCTATGGTGGAGGATCTCACAGCGGCTAGGAGCTCTCC 218
 Db 63672 GAGCGCGTGGGGGAGCTTCTATCCATGGTGTGAAGCGTGGAGCGGCTAGGGAACCTCC 63613
 Qy 219 TTCCCTAGCCGGATGACCTAACCGCGGTGCGCACCGCTTGGCGCGGCTCTCTGGGCGGT 278
 Db 63612 TTCCC-CGCCAGATGGAAGTCGATCAGTCGCGCTATTTGGCGGGGCTGTCTTCCCT 63554
 Qy 279 CCGGTGCAGCCTCTGTTCGGAAGCCCGCTCTCTGGGCTCTCTCTGCGGCGCGGAA 338
 Db 63553 GTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGTGGCTTGAAGA 63494
 Qy 339 TCGGCTGCGAGTACCCACTCGGTGGCTGGGAGGCGGAGACTGTGTAGACCTCGGATCC 398
 Db 63493 TCGGCTTGGAGCGAGCGCACCGCTGGGCAAGGCGGAGACTGTGTAGGCTTCTCCGA 63434
 Qy 399 AGCTTGGCTGACCGCGCTGAGCTCTCTCTCTCTCTCTGTCTGAGAAGCCGCAAGGA-AA 457
 Db 63433 ATCCGCTGCACTTCCAGCCCTGAGCGCGCGGCGCTTACCTGAGAGACTGTCAAGAAAA 63374
 Qy 458 GGAGATGAGACAGGAACGACAGGGCCCGGAAGCGGCTTGGGCTCTGGCTGGGCTCTGTG 517
 Db 63373 GGAGATGAGCGCGGAGACAGCGGATGCGGGAACCGCTTGGGCTCTGGGCGGCTTTCCG 63314
 Qy 518 GTTCCGCTGCGCC---TTCTTTCGGGATCGCAGCGCTCTCTTGGAGTTTCCCGCGCC 574
 Db 63313 GTTCTGGCCACCTTTTCTCTCGCGATCGCAAGCAGGCTCTTCTAAGTTTCCCGAGCC 63254
 Qy 575 TTCTCTCTCGACAAATCCCGGAA-----CTCGCTCTGCGCGGAGCGTCG 619
 Db 63253 TCTTGGCCCGGAAATCTCCGGAAACCCACACTGCTTCTCTGCGCAGCCGAGACTCG 63194
 Qy 620 GACGAGTACTGAGACCAATTTGTTTCTCAGTCTCTTGGCTCTCTCTCTAGCTATTTCCA 679
 Db 63193 GGTCACTTCTGAGCAAACTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTTGCTTCA 63134
 Qy 680 GAAGCTGCTCTTGGGAGCAGCTTTCGCGGCGCTGATTCCTACAGATGGCTAGATT 739
 Db 63133 GAAGGTCTTAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACCAAGTGGCTAGATT 63074
 Qy 740 TGGCGCAAGTTACAGCGCCCTGAGAGCTTTCGAGAGGACGGGAGGAATCTGACGCTCCAC 799
 Db 63073 TGCTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGGAGCGGAGAAACAGCGCCGCCAC 63014
 Qy 800 GGTGAGAGTCTCTGAGTTACTACTGCGGTGAGTCTTTCGCGAAGACTCGCGTCTGAG 859
 Db 63013 AGCGCAGAAATCTTTGAGTT-CGCTGAGCTCGACTCTCTC---AGACCCCTCGGTACCA 62958
 Qy 860 TACTCTTGCATTTGGCTAGAGAGGAGTCCAGTGGCAGTCTGCTCTCAGA-CTGGAAG 918
 Db 62957 GTCCCTTGTATGGCTAGAGAGGAGTCCACTGGCAATACTTCGCCCGGAGCTAAAT 62898
 Qy 919 TTAACCTCAAGGCGCCAGGAAAGAGCTTTAGACTCTGAGCGGCCCACTTTCTCTCTGGAGC 978
 Db 62897 TGGAGCTTAAGGCCCAAGGAAAGTCTTTTGGACCTGCGAGCACAGGCTTTTCTCTTAGAGC 62838
 Qy 979 AGCAGCTGTGGGAGTGGAGTGTGCTGCCAGTAGCTTCAAGCTGTGTAGTCTCCACCC 1038
 Db 62837 AGCAGCTGTGGGAGTGGAGCTGTGTCGAGTGTGCTTCAATCCCGTCTGTACTCTAACCC 62778
 Qy 1039 GAGAACTTGAATCTTTCATCTCTGGGCTCTGAGGCTTTCAGAGCTTAGGTAAATTTCAAGG 1098
 Db 62777 GGGAACTTGGCTCTTCCCTCTGGGCTCTTAACATTTCAACGATAGACNAATTTCAAGTG 62718

Qy 1099 TAGTTTCCTATCTCCTGAAACCCCTTCTCTACCTGGAGTACCTTTCCCAAGTTTAGGGCTGGCT 1158
Db 62717 TGGTATCCTATTTTCTGCTGAACCCCTTCTACCTGGAGTCTGTTTCTCTAGCTAGAGTCAAGCT 62658
Qy 1159 GTACAGCAGCGCTGGAGGTGGCCAGTTTGTGGGTTTCCGAACTAAACCCAGAGAGCT 1218
Db 62657 ATCAGAACAGTGATGGAATAGCAGAGGTAGTCGGCTTCCAGACACTAAACCCAGAGAGCA 62598
Qy 1219 GCTATCTTTCTGAAGATGGTTGTCAACCTCAGCCGTTTGGGGCAGAGATGTCGGCAACCG 1278
Db 62597 GCTGCCCTGAGAGAGGACCATTTGTCAATCCAGCGCTGAGTGCAAGAACTCATTCGCGCCT 62538
Qy 1279 CTTGGAGAGGTGTCGCGCTCTCTACAGAGAGCCCTGCGGAAATCCACACCTTCGTA 1338
Db 62537 CGTGGCAGGAGTGTCCACCTCTTTCTACGGAGGCTTACCAGAAATTCACCATCTCCGCA 62478
Qy 1339 TGAACCGCTAGATTTCTCT--CCAGGCTAACAAAGGCGCAAGGTATACCAACCCCTGCACC 1395
Db 62477 TGAACCGCTGGAATTTCTTCAACAGGCTAACAAAGGCGCAAGATTTACCAACCCCTGCACC 62418
Qy 1396 AAGATAATGGCTATCATAGCTCGGAGGAGGAACAATAACCTTCTCCGGATGACCCCAAC 1455
Db 62417 AGGATAATGGCTACCAAGCTCGGAGGAGGAACAACAGCTTCTCCGGATGATGCCAAAC 62358
Qy 1456 ATTGCACAGATAACCCAGCAGCAGCGGTGTCCTCTGTCGACAGCAGGCC----- 1504
Db 62357 ACTGCAGAGATAACCCCAACAGTGTGTTCTGCTGTCGAGAGATTTCTGGAAACACCC 62298
Qy 1505 -GGAGCCACTGAGAAACCAACCAATTTGGTATTCAAGAGTTTCAC----- 1551
Db 62297 AGGAATCCACTGAAAGAAATAAGATTTATTAATCTACAGAGTTTCCACTTGTCTTGGAAAG 62238
Qy 1552 ---AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAATTTACCGTGGGAAAGAAATGTGAAG 1608
Db 62237 AAGAGAGCCCTTCTGAGGCGTGTCCATCTAGTGAGATACCTATGGAAGAGGCGCTGGAG 62178
Qy 1609 AGGACCACTAATGCACTGACCTCTCAGATAGAGAGAGAGCCTTCTCTGTTTCTACCA 1668
Db 62177 AGGCGCAATTAAGTGTAGTTGATTACTCATACCTTAGAAGGTGACCTTTCCCATTTCTGCCA 62118
Qy 1669 GACCAGTTTGTAGCAACAACTGATAGATTATTTTGGAGGCGCCCGCCAGTACATTGG 1728
Db 62117 GACCAGCTTGTAGTAACAACTGATAGATTATTTTGGAGGCTGCAATCCAGTACCTGG 62058
Qy 1729 AAGCAGCTCTGATTTCTGAAAGTGAAGATTGGGCGGAGGAACCTTGAGGACGATGGCTTTG 1788
Db 62057 AAACAAGTTCTGATCCAGAGGTGAGGATTTGGGATGAGAGCTGAGATGATGGTTTG 61998
Qy 1789 ATAGCGATGGCTCCTGTCTGAATCAGACGTGGGAACAGGACTCGGAAGGCTTCCACCTTT 1848
Db 61997 ATAGTGATAGCTCACTGTCACTCAGACTCAGACCTTGAACAAGACCCCTGAAGGGCTTCCACCTTT 61938
Qy 1849 GGAATCTTTTCCAGTGTAGATCTTTACAAACCCCAAACTTTACAGCCACGATTCAGA 1908
Db 61937 GGAATCTTTTCTGAGTGTAGATCTTTATATATCCCAAGAACTTTACAGCAACAATTCAGA 61878
Qy 1909 CGGCTGCCAGAAATTTGCCCCAGAGACCATCAGATTTAGGAGACATCTCTGGTCTGGCAGCT 1968
Db 61877 CTGCTGCCAGAAATTTGTTCTGTAAGAGCTTCTGATTCAGAGAAGATTTGTCTGGCAAGT 61818
Qy 1969 GTGTGTAGGGA---GCTGTGAGAGGGAACCCCTTCCGGAGACCCCGACCATAGTTCCG 2025
Db 61817 CTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTG 61758
Qy 2026 GGGAGGAGATGACTGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTAATTTGGAACT 2085
Db 61757 GGGAGGAGATGACTGGGAATCTAGTGAGATGAAGCAGAGAGTCTCAAACTGTGGAACT 61698
Qy 2086 CTTTCTGTCTTCTGAGAGACCCCTACACCTTTTAAATTTTAAAGGCTCTTTTCAACCGT 2145
Db 61697 CATTTCTGTAATTTCTGATGACCCCTACAAACCTTTTAAATTTTAAAGGCTCTTTTCAACAT 61638
Qy 2146 CAGGGAAGAAATTGGAAGGCGCTCAGGACTCAAAAGGCGCTTCTCTGAGGTCAAGTGGCCT 2205

Db 61637 CAGGGGAAAATGAGAAGGCTGCTGACTCAAAGACCCCATCTGAGTCCATTGTGGCCA 61578
Qy 2206 TCTCTGGCCATCATACCTTACTTCTTTTAAGGCCAGCTGTTTAGAGAGCCAAAGAGATA 2265
Db 61577 TTTCTGAGTCTCACACTTACTTCTTTTAAGGTGACGCTGTTGGGGAGCCAAAGAGTG 61518
Qy 2266 ATTGTCCAGGCTGTGGCTGGGTGGGTGAGGCTTCTGCTGGAGAAAGATACACCCATATCAAGA 2325
Db 61517 AATGTCCAGACTCGGTACAGCGTGAAGCTTCTTTCTGGAGGAAGACACACATGTCAAAA 61458
Qy 2326 GAAAAAGGTAACTTCCTG 2345
Db 61457 GAAAAAGGTTTGTCTTCTG 61438
RESULT 14
AL627202/c
LOCUS AL627202 328863 bp DNA linear HTG 24-OCT-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-174P23, 2 unordered pieces.
ACCESSION AL627202
VERSION AL627202.6 GI:24394911
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 328863)
AUTHORS Clark, S.
TITLE Direct Submission
JOURNAL Submitted (21-Oct-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:20268846.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA174P23
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 170058 bases at least Q40
Consensus quality: 171825 bases at least Q30
Consensus quality: 172641 bases at least Q20
Insert size: 328763; sum-of-contigs
Insert size: 172230; 12.6% error; agarose-fp
Quality coverage: 2.67x in Q20 bases; sum-of-contigs Quality
coverage: 5.15x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6943: contig of 6943 bp in length
* 6944 7043: gap of 100 bp
* 7044 328863: contig of 321820 bp in length.
FEATURES
source
1. 328863
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-174P23"
/clone_lib="RPCI-11.1"
misc_feature 1. 6943


```

/note="assembly fragment: 03558"

```

QY 2026 GGGAGGAAGATGACTTGGGACCGAGTGCAGATGAAGCAGAGAGATCTTAAATTGTGGAAC 2085


```
Db 223544 GGGAGGAAGTCTAGTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGAACT 223485
Qy 2086 CTTTCTCTGCTTCTGAGGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTTTCAACCGT 2145
Db 223484 CATTCGTGTAATTCGTATGACCCCTACAACTTTTAAATTTTAAAGGCTCCTTTTCAACAT 223425
Qy 2146 CAGGGAAGAAATTGAAAGGCGCTCAGGACTCAAAAGGCTCTTCTTGAGGTGTCAGTGGCCT 2205
Db 223424 CAGGGAAGAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTGTGGCCA 223365
Qy 2206 TCTCTGCCATCATACCTTACTTCTTGTAAAGGCCAGCTGTGTAGAGAGCCCAAGAATA 2265
Db 223364 TTTCTGAGTGTCAACCTTACTTCTTGTAAAGTGTGACCTGTGTGGGAGGCCAAGAAAGTG 223305
Qy 2266 ATTGTCCAGGCTGTGGCTGGGTGAGGCTCTTGTCTGAGAAAGATACACCATATCAAGA 2325
Db 223304 ATGTCCAGACTCGGTACAGGTGACGTTCTTCTGAGGAGACACACATGTCAAAA 223245
Qy 2326 GAAAAAGGTAACTTCTCTG 2345
Db 223244 GAAAAAGGTTGTTCTTG 223225

RESULT 15
AC048369 166233 bp DNA linear HTG 25-JUN-2000
LOCUS Homo sapiens chromosome 1 clone RP11-284G5 map 1, WORKING DRAFT
DEFINITION SEQUENCE, 29 unordered pieces.
ACCESSION AC048369
VERSION AC048369.2 GI:8705131
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 1, clone RP11-284G5
Unpublished
2 (bases 1 to 166233)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McDrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,W., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7549699.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
```

```
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8045
Center clone name: 284_G_5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151910 bases at least Q40
Consensus quality: 158917 bases at least Q30
Consensus quality: 161704 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 163433; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1069: contig of 1069 bp in length
* 1070 1169: gap of 100 bp
* 1170 2738: contig of 1569 bp in length
* 2739 2838: gap of 100 bp
* 2839 4143: contig of 1305 bp in length
* 4144 4243: gap of 100 bp
* 4244 5489: contig of 1246 bp in length
* 5490 5589: gap of 100 bp
* 5590 6874: contig of 1284 bp in length
* 6875 9153: contig of 2180 bp in length
* 9154 9253: gap of 100 bp
* 9254 10555: contig of 1302 bp in length
* 10556 10655: gap of 100 bp
* 10656 13995: contig of 3340 bp in length
* 13996 14095: gap of 100 bp
* 14096 16689: contig of 2594 bp in length
* 16690 16789: gap of 100 bp
* 16790 19295: contig of 2506 bp in length
* 19296 19395: gap of 100 bp
* 19396 22797: contig of 3402 bp in length
* 22798 26453: contig of 3556 bp in length
* 26454 26553: gap of 100 bp
* 26554 31369: contig of 4816 bp in length
* 31370 31469: gap of 100 bp
* 31470 35528: contig of 4059 bp in length
* 35529 35628: gap of 100 bp
* 35629 40531: contig of 4903 bp in length
* 40532 40631: gap of 100 bp
* 40632 45154: contig of 4523 bp in length
* 45155 50400: contig of 5146 bp in length
* 50401 50500: gap of 100 bp
* 50501 56357: contig of 5857 bp in length
* 56358 56457: gap of 100 bp
* 56458 62667: contig of 6210 bp in length
* 62668 62767: gap of 100 bp
* 62768 69282: contig of 6515 bp in length
* 69283 76270: contig of 6888 bp in length
* 76271 76370: gap of 100 bp
* 76371 84430: contig of 8060 bp in length
* 84431 84531: gap of 100 bp
* 84531 92060: contig of 7530 bp in length
* 92061 92160: gap of 100 bp
* 92161 101111: contig of 8951 bp in length
* 101112 101212: gap of 100 bp
* 101213 109009: contig of 7798 bp in length
```


Db	3735	AGCCTGGAGAGGCGCGAATAAGTGTAGTTGATTACTCATACCTAGAAGGTGACCTTCCCA	3794
Qy	1660	TTTCTACAGACCAAGTTTGTAGCAAACTGATAGATTATATTTGGGAGGCGCCCCCA	1719
Db	3795	TTTCTGCCAGACCAAGCTTGTAGTAACAACCTGATAGATTATATTTGGGAGGTGCATCCA	3854
Qy	1720	GTGACTTGGAAAGCCAGCTCTGATTCTGAAAGTGAAGGATTGGGGCGAGGAACCTGAGGACG	1779
Db	3855	GTGACCTGGAAACAAGTTCTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATG	3914
Qy	1780	ATGCTTTGATAGCGATGGCTCCCTGCTGTAATCAGACGTGGAACAGGACTCGGAAGGCC	1839
Db	3915	ATGGTTTGTATGATGATAGCTCACTGTGCAGACTCAGACCTTGAACAAGACCCCTGAAGGGC	3974
Qy	1840	TTACACCTTTGGAACCTCTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTACAGCCA	1899
Db	3975	TTACACCTTTGGAACCTCTTCCGAGTGTAGATCCTTATATCCCGAATCTTTACAGCAA	4034
Qy	1900	CGATTTCAGACGGCTGCCAGAAATGCCCCACAGACCCCATCAGATTCAGGGACATCCTGGT	1959
Db	4035	CAATTTCAGACTGCTGCCAGAAATTTCTCTGAAGAGCCTTCTGATTCAGAGAAGGATTGT	4094
Qy	1960	CTGCGAGCTGTGGTGTAGGGA	1980
Db	4095	CTGGCAAGTCTGATCTAGAGA	4115

Search completed: September 16, 2005, 01:58:03
Job time : 22799.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:08:59 ; Search time 2648.83 Seconds
(without alignment)
12220.183 Million cell updates/sec

Title: US-10-650-482-3

Perfect score: 5468

Sequence: 1 cggctcctcgtcgtcgcctg.....attaaaaaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

N_Geneseq_16Dec04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002as:*
- 7: Geneseq2002bs:*
- 8: Geneseq2003as:*
- 9: Geneseq2003bs:*
- 10: Geneseq2003cs:*
- 11: Geneseq2003ds:*
- 12: Geneseq2004as:*
- 13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5468	100.0	5468	12	ADQ88459 Mouse GAD
2	1324	24.2	2942	12	ADQ88457 Human GAD
3	1320.8	24.2	2942	4	AH18699 Human cDN
4	1189.2	21.7	2324	3	AAH98158 Human col
5	542.6	9.9	883	4	AAH99389 Human pro
6	239.2	4.4	772	4	AAH06623 Human cDN
7	181.6	3.3	540	6	ABK63389 Rat seque
8	181.6	3.3	540	10	ADBS57750
9	170.8	3.1	399	8	ABX52167 Bovine ES
10	113.6	2.1	689	4	AAI96694 Human neu
11	109	2.0	421	9	ACH49151 Human leu
12	109	2.0	543	4	AAH11666 Human cDN
13	99.6	1.8	1028	3	AAF21761 Human bre
14	98	1.8	482	4	AAH13104 Human cDN
15	98	1.8	2148	4	AAH16576 Human cDN
16	98	1.8	2148	6	ABV77954 Hypoxia-r
17	98	1.8	2148	10	ADG32888 Human DNA
18	93.8	1.7	218	3	AAH29803 Human sec
19	85.8	1.6	10610	12	ADQ97608 Mouse can
20	84	1.5	1678	2	AAT34611 Probe for

c	21	82.6	1.5	90141	12	ADQ97867	Adq97867 Mouse can
c	22	79.4	1.5	8351	6	ABK63492	Abk63492 Rat seque
c	23	79.4	1.5	8351	10	ABT41759	Abt41759 Toxicity
c	24	79.4	1.5	8351	12	ADP72626	Adp72626 Renal tox
c	25	79.4	1.5	119414	11	ACN45052	Acn45052 Mouse gen
c	26	79.2	1.4	1448	10	ADBS8896	Adbs8896 Toxicity-
c	27	79.2	1.4	1448	10	ADBS3612	Adbs3612 Primary r
c	28	79.2	1.4	1448	10	ADBS3612	Adbs3612 Toxicity
c	29	79.2	1.4	1454	3	AAC97642	Aac97642 Rat Bridg
c	30	79.2	1.4	4529	10	ADBS8829	Adbs8829 Toxicity-
c	31	79.2	1.4	4529	10	ADBS3548	Adbs3548 Primary r
c	32	79.2	1.4	4529	10	ABT42350	Abt42350 Toxicity
c	33	79.2	1.4	25891	11	ACN44952	Acn44952 Mouse gen
c	34	79.2	1.4	53963	9	ADA02537	Ada02537 Mouse Tco
c	35	79.2	1.4	53963	10	ADBS72275	Adbs72275 Mouse Tco
c	36	79.2	1.4	53963	10	ADBS95785	Adbs95785 Mouse Tco
c	37	78.8	1.4	3046	2	AAV29074	Aav29074 Rattus no
c	38	78.6	1.4	1496	6	ABL34990	Ab134990 Rat cDNA
c	39	78.6	1.4	13894	10	ADBS9134	Adbs9134 Toxicity-
c	40	78.2	1.4	71132	11	ACN45092	Acn45092 Mouse gen
c	41	77.8	1.4	1432	10	ADBS8133	Adbs8133 Toxicity-
c	42	77.8	1.4	1432	10	ADBS2618	Adbs2618 Primary r
c	43	77.6	1.4	5531	10	ADBS3403	Adbs3403 Primary r
c	44	77.6	1.4	5531	10	ABT42272	Abt42272 Toxicity
c	45	77.6	1.4	215980	6	AAH38337	Aal38337 Complemen
c	46	77.2	1.4	433	12	ADI04042	Adi04042 Human neu
c	47	77.2	1.4	233060	11	ACN43912	Acn43912 Mouse gen
c	48	77	1.4	36119	13	ABD33605	Abd33605 Murine ca
c	49	76.6	1.4	6251	6	ABK62565	Abk62565 Rat seque
c	50	76.6	1.4	6251	10	ABT40825	Abt40825 Toxicity
c	51	76.6	1.4	6251	12	ADP71847	Adp71847 Renal tox
c	52	76	1.4	478	10	ADBS1015	Adbs1015 Primary r
c	53	76	1.4	479	10	ADBS1404	Adbs1404 Primary r
c	54	76	1.4	575	10	ADBS7185	Adbs7185 Toxicity-
c	55	76	1.4	120239	13	ABD33244	Abd33244 Murine ca
c	56	75.4	1.4	508	6	ABK62412	Abk62412 Rat seque
c	57	75.2	1.4	2070	9	ACF36089	Acf36089 Rat FXR p
c	58	75.2	1.4	2070	9	ACF04415	Acf04415 Rat farne
c	59	75.2	1.4	2070	10	ADBS8407	Adbs8407 Toxicity-
c	60	75.2	1.4	2070	10	ADBS52987	Adbs52987 Primary r
c	61	75.2	1.4	2070	10	ABT42054	Abt42054 Toxicity
c	62	75.2	1.4	2070	12	ADP72840	Adp72840 Renal tox
c	63	75.2	1.4	3122	5	AAH1297	Aaa91297 Rat Urote
c	64	75.2	1.4	3122	13	ADM99718	Adm99718 Norway ra
c	65	75	1.4	2472	10	ADBS8312	Adbs8312 Toxicity-
c	66	75	1.4	2472	10	ADBS2866	Adbs2866 Primary r
c	67	75	1.4	2472	12	ADJ10118	Adj10118 Rat vitam
c	68	75	1.4	2485	10	ADBS3252	Adbs3252 Primary r
c	69	75	1.4	2485	10	ADBS42199	Adbs42199 Toxicity
c	70	74.8	1.4	2010	6	AAH94721	Aaa94721 Rat seque
c	71	74.6	1.4	227	5	AAH82230	Aah82230 Rat diffe
c	72	74.4	1.4	10929	13	ADT66569	Adt66569 Rat lutzp
c	73	74.2	1.4	13011	2	AAT96631	Aat96631 cDNA enco
c	74	74.2	1.4	13011	10	ABT42448	Abt42448 Toxicity
c	75	74.2	1.4	13011	12	ADP72914	Adp72914 Renal tox
c	76	74	1.4	411	12	ADP72375	Adp72375 Renal tox
c	77	74	1.4	507	12	ADP71796	Adp71796 Renal tox
c	78	74	1.4	3134	6	ABL34973	Ab134973 Murine cD
c	79	73.6	1.3	1125	12	ADQ35602	Adq35602 Novel mou
c	80	73.6	1.3	7577	13	ABD32996	Abd32996 Mouse can
c	81	73.6	1.3	7997	2	AAQ89553	Aaq89553 Rat chole
c	82	73.6	1.3	7997	2	AAT79737	Aat79737 Rat chole
c	83	73.6	1.3	109147	12	ADQ97718	Adq97718 Mouse can
c	84	73.4	1.3	427	2	AAT34599	Aat34599 Probe for
c	85	73	1.3	38239	12	ADQ97626	Adq97626 Mouse can
c	86	73	1.3	109565	13	ABD33086	Abd33086 Murine ca
c	87	72.8	1.3	1089	10	ADCS3443	Adcc3443 Rat pl8AB
c	88	72.8	1.3	1665	3	AAA90869	Aaa90869 Rat stres
c	89	72.8	1.3	1686	10	ADBS3061	Adbs3061 Primary r
c	90	72.8	1.3	2006	10	ADBS8134	Adbs8134 Toxicity-
c	91	72.8	1.3	2006	10	ADBS2619	Adbs2619 Primary r
c	92	72.8	1.3	2006	10	ABT41851	Abt41851 Toxicity
c	93	72.8	1.3	2279	10	ABR56187	Abd56187 Toxicity-

	72.8	1.3	4964	12	ADP72921	Renal tox
c 94	72.8	1.3	4964	12	ADP72921	Adp72921
c 95	72.6	1.3	44576	3	AZ61522	Az61522
c 96	72.6	1.3	507538	11	ACN45088	Mouse gen
c 97	72.6	1.3	231004	12	ADQ97855	Mouse can
c 98	72.4	1.3	864	4	ABA77056	Proliferase
c 99	72.4	1.3	4039	12	ADJ35834	Rat proop
c 100	72.4	1.3	189158	11	ACN441124	Mouse gen

ALIGNMENTS

RESULT 1

AD088459

ID ADQ88459 standard; cDNA; 5468 BP.

AC ADQ88459:

07-OCT-2004 (first entry)

XX
DE
Mouse GADD34-like (GADD34L) cDNA.

AA
KW Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;
KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;
KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; mouse; gene;

XX

[illegible]

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	CDS	FT	CDS
462.	2558	462.	2558

```
FT /*tag= a
```

FT
/product= "Mouse GADD34-like (GADD34L) protein"

FT /transl_except= (pos:1332. .1334, aa:Arg)

```
FT
/ transl_except= (pos:1341. .1343, aa:Arg)
```

```
FT
/transl_except= (pos:1344. .1346, aa:Trp)
```

```
FT /transl_except= (pos:1350. .1352, aa:Val)
```

```
FT      /transl_except= (pos:1362. .1364, aa:Pro)
```

```
FT      /transl_except= (pos:1368. .1370, aa:Gln)
```

```
FT      /transl_except= (pos:1377. .1379, aa:Asp)
am      /transl_except= (pos:1380. .1381, aa:Asp)
```

```
FT
vv
/transl_except= (pos:1389. .1391, aa:Leu)
```

XX
PN IIS2004142345-A1

XX
PD 22-111.-2004

28-AUG-2003 2003US-00650482

06-SEP-2002. 2003US-0408679D

XX
DA (POND /) BOW D

PA (ROND/) RON D.
PA (TOIS/) TOISSE C

XX
PT
Bon D
Toussaint C.

XX	WT	CCC
CC	CCC	CCC

DR WPI; 2004-552556/
DR R-REDB: APO88460

xx Screening test substances for preventing or treating disease involving oxidative stress, by testing test substances for its ability to inhibit activity of GADD34L and identifying test substance that inhibits activity of GADD34L.

PS Disclosure: SEO ID NO 3: 30pp: English.

The present invention relates to a method of screening several test substances for preventing or treating diseases involving oxidative stress such as neuronal ischaemia, heart ischaemia, renal damage induced by ischaemia or toxins, autoimmune diseases and neurodegenerative disorders. The method involves testing the test substances for its ability to inhibit the activity of GADD34-like (GADD34L), also referred to as eIF2alpha-specific regulatory subunit of phosphatase, and identifying the test substance which inhibits the activity of GADD34L. The present

CC	sequence	is mouse	GADD34L	cDNA.
XX				
SQ	Sequence	5468 BP; 1321 A; 1272 C; 1366 G; 1509 T; 0 U; 0 Other;		
	Query Match	100.0%;	Score 5468;	DB 12; Length 5468;
	Best Local Similarity	100.0%;	Pred. No. 0;	
	Matches 5468;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	CGGTCTCTCCGTCTCGCCCTCAGCTTC	CGGGTGTGCGGCTGCGGCCATTTGAGCTTCGC	60
Db	1	CGGTCTCTCCGTCTCGCCCTCAGCTTC	CGGGTGTGCGGCTGCGGCCATTTGAGCTTCGC	60
Qy	61	TTCTTTTGGGCCCTCGCTCGCATCCAGCA	CCCTCTTCGCGCTTGGCGCTTCGCGGCTCCG	120
Db	61	TTCTTTTGGGCCCTCGCTCGCATCCAGCA	CCCTCTTCGCGCTTGGCGCTTCGCGGCTCCG	120
Qy	121	TGGGGGCCACCGGAACGCGCGTCTGCTCC	CGCGCGCGCGCGCGAGAGGCTCTTCT	180
Db	121	TGGGGGCCACCGGAACGCGCGTCTGCTCC	CGCGCGCGCGCGCGAGAGGCTCTTCT	180
Qy	181	CTATGTTGGAGCGATCTCACACCGGCTTAG	CAAGCTCTCTCCCTAGCCGGATGGACCTA	240
Db	181	CTATGTTGGAGCGATCTCACACCGGCTTAG	CAAGCTCTCTCCCTAGCCGGATGGACCTA	240
Qy	241	ACCCTGGTGCACACCGCTTGGCGGGCTCT	GGGCGCTTGGGCGAGCATCTGCTGCGGA	300
Db	241	ACCCTGGTGCACACCGCTTGGCGGGCTCT	GGGCGCTTGGGCGAGCATCTGCTGCGGA	300
Qy	301	AGCGCGGCTCTCTGGGCTCTCTTGGCGGCT	CGGGAATCGGACTGCACTCCACTCCG	360
Db	301	AGCGCGGCTCTCTGGGCTCTCTTGGCGGCT	CGGGAATCGGACTGCACTCCACTCCG	360
Qy	361	TGGCTGGCAAGCGGAGACTGTGTAGACTCT	CGGATCCAGCTCGCTGACGCGCTGAG	420
Db	361	TGGCTGGCAAGCGGAGACTGTGTAGACTCT	CGGATCCAGCTCGCTGACGCGCTGAG	420
Qy	421	CTCTGTCTCTCTCTGTCTGAGAACCGCGCA	AGGAAGAGATGGAGACAGGAACGACCA	480
Db	421	CTCTGTCTCTCTCTGTCTGAGAACCGCGCA	AGGAAGAGATGGAGACAGGAACGACCA	480
Qy	481	GGGCGCGGAAGCGGCTGGCCCTCGGCTGGG	CTCTGGTTCGGGCTGCGGCTTCTTTCGGC	540
Db	481	GGGCGCGGAAGCGGCTGGCCCTCGGCTGGG	CTCTGGTTCGGGCTGCGGCTTCTTTCGGC	540
Qy	541	GATCGACAGCTCTCTTTCGGAGTTCCCGCG	CGCTTCTCTCGACAAATCCCGGGAAC	600
Db	541	GATCGACAGCTCTCTTTCGGAGTTCCCGCG	CGCTTCTCTCGACAAATCCCGGGAAC	600
Qy	601	CCGCTCTGCGCGAGCGTCGACACAGGTACT	TGGAACCAATTTGCTTCTCAGCTCTTTCGCC	660
Db	601	CCGCTCTGCGCGAGCGTCGACACAGGTACT	TGGAACCAATTTGCTTCTCAGCTCTTTCGCC	660
Qy	661	TGCTCCCTAGCTATTTCAGAACTGCTGCTTT	TGGAGCCAGCTTTCGGGGGCTGATTC	720
Db	661	TGCTCCCTAGCTATTTCAGAACTGCTGCTTT	TGGAGCCAGCTTTCGGGGGCTGATTC	720
Qy	721	CTACCAAGTGGCTAGATTTTGGCGAAGTTT	TACAGCCCTTGAGAGCTTCGAGAGGACGGG	780
Db	721	CTACCAAGTGGCTAGATTTTGGCGAAGTTT	TACAGCCCTTGAGAGCTTCGAGAGGACGGG	780
Qy	781	AGGAATCTGACGCTCCACCGTTCGAGAAGT	CTCTGAGTTACATCGGGCTGGACTCTTCG	840
Db	781	AGGAATCTGACGCTCCACCGTTCGAGAAGT	CTCTGAGTTACATCGGGCTGGACTCTTCG	840
Qy	841	CGAAGACTCGCTGCTGATCTCTTTGCAATT	TGGCTAGAGAGGGAATCCAGTGGCAGTGC	900
Db	841	CGAAGACTCGCTGCTGATCTCTTTGCAATT	TGGCTAGAGAGGGAATCCAGTGGCAGTGC	900
Qy	901	TGCTCTCTCAGCTGGAAGTTAACTCAAGCC	CAGGAAGAGCTTTAGACTCTGACGCGC	960
Db	901	TGCTCTCTCAGCTGGAAGTTAACTCAAGCC	CAGGAAGAGCTTTAGACTCTGACGCGC	960
Qy	961	CCACTTTCTCTCTGGAGCAGCAGCTGTGGG	GAGTGAGTTGCTGCCAGTAGCTTCAAG	1020

Db ||||| 961 CCACTTTCTCTCGAGCAGAGCTGTGGGAGTGGAGTTGCTGCCAGTAGTCCCTCAAG 1020
Qy ||||| 1021 CTGCTAGTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGCGTTGAGA 1080
Db ||||| 1021 CTGCTAGTCTCCACCGAGAACTTGACTCTTCATCTCTGGGCTCTGAGCGTTGAGA 1080
Qy ||||| 1081 GCTTAGGTAAATTCAGAGTGTCTTCTATCTCTGAAACCTTCTTACCTGGAACCTTTC 1140
Db ||||| 1081 GCTTAGGTAAATTCAGAGTGTCTTCTATCTCTGAAACCTTCTTACCTGGAACCTTTC 1140
Qy ||||| 1141 CCCAGTTAGGCTGGCTGTCTGAGCAGCGCTGGAGTGGCAGTTTGTGGTTCCGAA 1200
Db ||||| 1141 CCCAGTTAGGCTGGCTGTCTGAGCAGCGCTGGAGTGGCAGTTTGTGGTTCCGAA 1200
Qy ||||| 1201 CACTAACCCAGAGAGCTGCTATCTTCTGAAGATGGTTGTTCACCTCAGCGCTGGGG 1260
Db ||||| 1201 CACTAACCCAGAGAGCTGCTATCTTCTGAAGATGGTTGTTCACCTCAGCGCTGGGG 1260
Qy ||||| 1261 CAGAGATGTGGCAACCGCTGGAGAAAGTGTCCGCTCTCTACAGAAAGGCTGGCGG 1320
Db ||||| 1261 CAGAGATGTGGCAACCGCTGGAGAAAGTGTCCGCTCTCTACAGAAAGGCTGGCGG 1320
Qy ||||| 1321 AAATCCACCACCTTCTGATGAAAAGGCTAGAAATTCCTCCAGGCTAAACAAAGGGCAAGT 1380
Db ||||| 1321 AAATCCACCACCTTCTGATGAAAAGGCTAGAAATTCCTCCAGGCTAAACAAAGGGCAAGT 1380
Qy ||||| 1381 TACCACCCCTGACCAAGATAATGGCTATCATAGCTTGGAGAGGAAACATAACCTTCTCC 1440
Db ||||| 1381 TACCACCCCTGACCAAGATAATGGCTATCATAGCTTGGAGAGGAAACATAACCTTCTCC 1440
Qy ||||| 1441 GGATGGACCCACAACTTGCA CAGATAACCCAGACAGGCGGTGTCCCTGCTGCAGACA 1500
Db ||||| 1441 GGATGGACCCACAACTTGCA CAGATAACCCAGACAGGCGGTGTCCCTGCTGCAGACA 1500
Qy ||||| 1501 GCGCGGACCCACCTGAGNAAAAACAGAAATTTGGTGATTCAGAAATTTTCAGAGCGCCC 1560
Db ||||| 1501 GCGCGGACCCACCTGAGNAAAAACAGAAATTTGGTGATTCAGAAATTTTCAGAGCGCCC 1560
Qy ||||| 1561 AGGGAAGCAGTCTGTTTGTGAAATPACCCGTGGAAAAAGATGTGAAGAGGACACACTA 1620
Db ||||| 1561 AGGGAAGCAGTCTGTTTGTGAAATPACCCGTGGAAAAAGATGTGAAGAGGACACACTA 1620
Qy ||||| 1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTTTCTACAGACCACTTTGTA 1680
Db ||||| 1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTTTCTACAGACCACTTTGTA 1680
Qy ||||| 1681 GCAACAACTGATAGATATATTTTGGAGGCGCCCGCAGTGACTTGGAAAGCCAGCTCTG 1740
Db ||||| 1681 GCAACAACTGATAGATATATATTTTGGAGGCGCCCGCAGTGACTTGGAAAGCCAGCTCTG 1740
Qy ||||| 1741 ATTCGAAAGTGAGATTTGGGCGGAGGAACTCTGAGGACGATGGCTTTTGATAGCGATGGCT 1800
Db ||||| 1741 ATTCGAAAGTGAGATTTGGGCGGAGGAACTCTGAGGACGATGGCTTTTGATAGCGATGGCT 1800
Qy ||||| 1801 CCCTGTCTGAACTCAGACGTGGAAACAGACTCGGAAGGCTTTCACCTTTGGAACTCTTTC 1860
Db ||||| 1801 CCCTGTCTGAACTCAGACGTGGAAACAGACTCGGAAGGCTTTCACCTTTGGAACTCTTTC 1860
Qy ||||| 1861 ACAGTGTAGATCTTTACAAACCCCAAACTTTACGCCACGATTCAGACGCTGCCAGAA 1920
Db ||||| 1861 ACAGTGTAGATCTTTACAAACCCCAAACTTTACGCCACGATTCAGACGCTGCCAGAA 1920
Qy ||||| 1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGGTCTGGCAGCTGTGGTGTAGGGA 1980
Db ||||| 1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGGTCTGGCAGCTGTGGTGTAGGGA 1980
Qy ||||| 1981 GCTGTCCAGAGGACCCCTTCCGAGACCCCGACCATAGTTCCGGGAGGAAAGTGAAT 2040
Db ||||| 1981 GCTGTCCAGAGGACCCCTTCCGAGACCCCGACCATAGTTCCGGGAGGAAAGTGAAT 2040
Qy ||||| 2041 GGGAAACGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAATCTTTTCTGTCAATCTG 2100

Db ||||| 2041 GGGAAACGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAATCTTTTCTGTCAATCTG 2100
Qy ||||| 2101 AGGACCCCTTACAAACCTTTTAAATTTTAAAGGCTCTCTTTTCAACCGTCAAGGAAGAAATTTGA 2160
Db ||||| 2101 AGGACCCCTTACAAACCTTTTAAATTTTAAAGGCTCTCTTTTCAACCGTCAAGGAAGAAATTTGA 2160
Qy ||||| 2161 AGGCGCTCAGGACTCAAGGCGCTCTTCTGAGGTCAAGTGGCCTTCTCTGGCCATCATTA 2220
Db ||||| 2161 AGGCGCTCAGGACTCAAGGCGCTCTTCTGAGGTCAAGTGGCCTTCTCTGGCCATCATTA 2220
Qy ||||| 2221 CCTTACTTCTCTTAAGGCCACGTGTTAGAGCCAAAGAGATAATTTGTCCAGGCTGTG 2280
Db ||||| 2221 CCTTACTTCTCTTAAGGCCACGTGTTAGAGCCAAAGAGATAATTTGTCCAGGCTGTG 2280
Qy ||||| 2281 GGCTGGGTGAGGCTCTTTGCTGGAGAAAGATACACCCATATCAAGAGAAAAAAGTAACTT 2340
Db ||||| 2281 GGCTGGGTGAGGCTCTTTGCTGGAGAAAGATACACCCATATCAAGAGAAAAAAGTAACTT 2340
Qy ||||| 2341 TCCTGGAAGAAGTTACTGAGTATTTATTAAGTGGTGATGAGGATCGCAAAAGGACCATGGG 2400
Db ||||| 2341 TCCTGGAAGAAGTTACTGAGTATTTATTAAGTGGTGATGAGGATCGCAAAAGGACCATGGG 2400
Qy ||||| 2401 AAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAAAAGAAATTTCAAGAAAAAGAAAGTTGCCA 2460
Db ||||| 2401 AAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAAAAGAAATTTCAAGAAAAAGAAAGTTGCCA 2460
Qy ||||| 2461 TTGGCTACTCTTTGGCCCTTTGAGCACAGAGAAAAAATGTTTAAATAGACTGAGGATCGAGT 2520
Db ||||| 2461 TTGGCTACTCTTTGGCCCTTTGAGCACAGAGAAAAAATGTTTAAATAGACTGAGGATCGAGT 2520
Qy ||||| 2521 CAAAGGACTTACTGTTGTA CAGCAATGTTAAGAGTGAAACAGCCTGCAACCCGTGCCAC 2580
Db ||||| 2521 CAAAGGACTTACTGTTGTA CAGCAATGTTAAGAGTGAAACAGCCTGCAACCCGTGCCAC 2580
Qy ||||| 2581 TCTGTCTCTTACTTGAGGTTTCCCTTAAAAACAAACACCTGGCAGCTGCTCTTGGACATG 2640
Db ||||| 2581 TCTGTCTCTTACTTGAGGTTTCCCTTAAAAACAAACACCTGGCAGCTGCTCTTGGACATG 2640
Qy ||||| 2641 TTTTAAAGAAAAACAACTTGTATCTAGAGATGCAAGTTTGAATTTTGGGTAAATGTCTCT 2700
Db ||||| 2641 TTTTAAAGAAAAACAACTTGTATCTAGAGATGCAAGTTTGAATTTTGGGTAAATGTCTCT 2700
Qy ||||| 2701 CATTAGAAAAACCAACTCCGATTAATGAAGAAATCTTCTATCTGTAATCCTCTCTTTTCCCTA 2760
Db ||||| 2701 CATTAGAAAAACCAACTCCGATTAATGAAGAAATCTTCTATCTGTAATCCTCTCTTTTCCCTA 2760
Qy ||||| 2761 TTTTAGTTGGATGTGGGTTTGTCTTTTGGAGGGCTCTCACTGCATAAATTTCTGTTTGGCC 2820
Db ||||| 2761 TTTTAGTTGGATGTGGGTTTGTCTTTTGGAGGGCTCTCACTGCATAAATTTCTGTTTGGCC 2820
Qy ||||| 2821 TAGGCTCTATGTATGTAGACCAAGTTGGCCTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880
Db ||||| 2821 TAGGCTCTATGTATGTAGACCAAGTTGGCCTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880
Qy ||||| 2881 TGTGCCACCAATGCCAGCTGGAAAGTGTGTTTTTAAATATCTCTGCTTACCTGGGGTGAGAG 2940
Db ||||| 2881 TGTGCCACCAATGCCAGCTGGAAAGTGTGTTTTTAAATATCTCTGCTTACCTGGGGTGAGAG 2940
Qy ||||| 2941 TTTTAAATTTTGCATCTTTCANAAACGTTTCTTGGAGGCGAGGGCTGGTCTGTTTTTACATAC 3000
Db ||||| 2941 TTTTAAATTTTGCATCTTTCANAAACGTTTCTTGGAGGCGAGGGCTGGTCTGTTTTTACATAC 3000
Qy ||||| 3001 AGGCTTCAGGCCAGCTAGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAAACAGAAACAA 3060
Db ||||| 3001 AGGCTTCAGGCCAGCTAGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAAACAGAAACAA 3060
Qy ||||| 3061 AACAGTCAGGTTACTGTGGAAA CTGAGGACAGGAGATAGGAAGGTCCAGGTATGCTTGGGA 3120
Db ||||| 3061 AACAGTCAGGTTACTGTGGAAA CTGAGGACAGGAGATAGGAAGGTCCAGGTATGCTTGGGA 3120
Qy ||||| 3121 ACTCAGTGAATCTCTGAACCTCAAGCCAGCATGGGGAGTTTAGCAAGACCTTTTACACTC 3180
Db ||||| 3121 ACTCAGTGAATCTCTGAACCTCAAGCCAGCATGGGGAGTTTAGCAAGACCTTTTACACTC 3180


```

Db 5341 CTTTTTCAGCTCAGTCATTTTACGTGTAACTGGAGATTTGATAGCTCAGATTATTTGT 5400
Qy 5401 ATATAATTTAATACTAATCTGTAATAATTGTAATAAATATATTTGCAATATTAATAAAAAA 5460
Db 5401 ATATAATTTAATACTAATCTGTAATAATTGTAATAAATATATTTGCAATATTAATAAAAAA 5460
Qy 5461 AAAAAAAA 5468
Db 5461 AAAAAAAA 5468

RESULT 2
ADQ88457
ID ADQ88457 standard; cDNA; 2942 BP.
AC ADQ88457;
XX
XX 07-OCT-2004 (first entry)
XX Human GADD34-like (GADD34L) cDNA.
DE
DE Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;
KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;
KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human; gene;
KW ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 407..2548
FT FT /*tag= a
FT FT /product= "Human GADD34-like (GADD34L) protein"
XX
XX US2004142345-A1.
XX
XX 22-JUL-2004.
XX
XX 28-AUG-2003; 2003US-00650482.
XX
XX 06-SEP-2002; 2002US-0408679P.
XX
XX (ROND/) RON D.
PA (JOUS/) JOUSSE C.
XX
XX Ron D, Jousse C;
XX
XX WPI; 2004-552556/53.
DR P-PSDB; ADQ88458.
DR GENBANK; AK027650.
XX
XX Screening test substances for preventing or treating disease involving
PT oxidative stress, by testing test substances for its ability to inhibit
PT activity of GADD34L and identifying test substance that inhibits activity
PT of GADD34L.
XX
XX Disclosure; SEQ ID NO 1; 30pp; English.
XX
XX The present invention relates to a method of screening several test
CC substances for preventing or treating diseases involving oxidative stress
CC such as neuronal ischaemia, heart ischaemia, renal damage induced by
CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.
CC The method involves testing the test substances for its ability to
CC inhibit the activity of GADD34-like (GADD34L), also referred to as
CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the
CC test substance which inhibits the activity of GADD34L. The present
CC sequence is human GADD34L cDNA.
XX
XX Sequence 2942 BP; 732 A; 739 C; 725 G; 746 T; 0 U; 0 Other;
XX
XX Query Match 24.2%; Score 1324; DB 12; Length 2942;
XX Best Local Similarity 72.5%; Pred. No. 0;
XX Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;

```

```

Qy 151 CCGTCCCGCGCGCGAGGGAGGGTTCTTCTATATGTGGAGCGATCTCACAGGCGCTAGG 210
Db 96 CCATCCCGAGGCGCGTGGGGGAGCTCTATATCCATGTGTGTGAAGCGTCGAGCGCGACTAGG 155
Qy 211 AGTCTCTCTTCCCTAGCCGCGGATGGACCTAAACCGCGGTGCGCACCGCTTTCGCGGGCGCTC 270
Db 156 GAACCTCCCTTCCC - CGCCAGGATGGAAATCGCATAGTCGCCGCCCTATTTCGCGGGCGTGT 214
Qy 271 TGGGCGGTCCGGTGCAGCACTCGTTGCGGAAGCGCGCTCTCTGCGGCTCTCTCTCCCG 330
Db 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGCTGG 274
Qy 331 CGCGGAATCGGATCGAGTACCACTCCGTGTGGTGGGCAAGCGGAGAGTGTGTAGACC 390
Db 275 CTCGAAGATCGGCTTGGAGCAGCGCACCGCTGCGCAAGCGCCGAGACTCTGTAGGCT 334
Qy 391 TGGGATCCAGCGCTGCGTGCAGCGCGCTGAGCTCTGCTCTCTCTCTCTGAGAAAGCGGCC 450
Db 335 TCCTCCGAATCCCGTGCAGCTCTCAGCGCGTGCAGCGCGCGGCCCTTACTCGAGAGACTGTC 394
Qy 451 AAG-GAAAGGAGATGGAGACAGGAACGCACAGGCGCGCGGAAGCGGCTTGGCCCTCGGCTG 509
Db 395 AAAAAAAGGAGATGGAGCGGGGACAGCGGATCGCGGAACCGCTTGGCCCTCGGGCG 454
Qy 510 GGCTCTCTGTTCGGCTGCGCC---TTCTTTGCGCGATCGCACGCTGCTCTTCGAGATTTC 566
Db 455 GGCTTCGGTTCGCGCACCCCTTTTCCCTCGCGGATCGCAAGCAGGCTCTTCTTAAGTTC 514
Qy 567 CGCGCGCTTCTCTCGACAAAATCCGGGAATCC-----GCTCTGCC 611
Db 515 CCAGCGCTTCTTGGCGCGGAAAATCTCCGGGAACCCACACTGCTTCTCTGCGCCAGCCC 574
Qy 612 GAGCTGCGGACAGGTACTGGACCAATTCCTTCTCAGCTCCTTGCCTGCTCCCTTAGC 671
Db 575 GAGACTCGGGTCAGTTACTGGACGAAACTGCTCTCCAGCTCTTGGGCGCTCCCGGA 634
Qy 672 CTATTCAGAAAGCTGCTGCTTTTGGAGCCAGCTTTTCCGGGGCGCTGATTCCTACAGATGG 731
Db 635 TTGCTTTCAGAAAGTGTCTAATTTGGAGCCAACTTTTCGGTGAATGTTTCCGACAGATGG 694
Qy 732 CTAGATTTTCCCGCAAGTTACAGCGCCCTGAGAGCTTCGAGAGAGCGGAGGATCTGAC 791
Db 695 CTAGATTTTCTGGAGTCTACAGCGCCCTGAGAGCCCTGAAGGCGCGGGAAGAACAGACC 754
Qy 792 GCTCCACAGGTGCAGAAAGTCTCTGAGTTACACTGCGGTGGAGCTCTTCGCGAAGACTCGC 851
Db 755 GCCCCACAGCGCAGAAATCTTTGAGTT -CGCTGCACTCGACTCCTC---AGACCCCTC 810
Qy 852 GTCGTCAGTACTCTTTCATTTGGCTAGAGAGGAGTCCAGTGGCAGTGTCTGCTCCTCAGA 911
Db 811 GGTCAACAGTCCCTTTGATTGGCTAGAGAGGGGATCCACTGGCAATACTCGGCCCCAGA 870
Qy 912 -CTGGAAGTTAAACTCAAGCCCGAGAAAGAGCTTTAGACTCTGCGAGCGGCCACTTTCCT 970
Db 871 CCTAAAATTTGAGACTTTAAGGCCAAGGGAAGTGTCTTTGGACCCCTGCGAGCAGAGCTTCT 930
Qy 971 CTTGAGGACGACGCTGTGGGGAGTGGAGTTGCTGCCAGTAGTCTTCAAGCTGTCTAGT 1030
Db 931 CTTAGAGCAGCAGCTGTGGGGAGTGGAGCTGTTGCCAGTAGCTTCAATCCCGTCTGTA 990
Qy 1031 CTCCACCGGAAACTTGGCTCTTTCATCTCTTGGGCGCTCTGAGCGTTTCAGAGCTTAGGTAA 1090
Db 991 CTCTAACCGGAACTTGGCTCTTTCGCCCTCTCGGCCCTCTAAACATTTCAACGCATAGACAA 1050
Qy 1091 TTTCAAGGTAGTTTCTCTATCTCTGAAACCCCTTCTACCTGGAGTACCTTCCCGAGTAGG 1150
Db 1051 TTTTCAGTGTGTATCTCTATTGCTGAACCCCTTCTACCTGGAGTGTCTTCTTAGGCTAGA 1110
Qy 1151 GCTGCGCTGTACAGCAGCGCTGAGGTGCGGAGTTCGCGGTTTTCGGAACACTTAACCCC 1210
Db 1111 AGTCAGCTATCAGAAACAGTGTGAANAATAGCGAGGTAGTCGGCTTCCAGACACTTAACCCC 1170

```

Qy	1211	AGAGAGCTGCTATCTTTCTGAAGATGGTTGTTCACCCCTCAGCCGTTGCGGCGAGAGATGTC	1270
Db	1171	AGAGAGCAGCTGCCTGAGAGAGGACCAATTTGTTCATCCCGAGCCGCTGAGTGCAGAACTCAT	1230
Qy	1271	GGCAACCGCTGGAGAAAGGTGTCGCGCTCTCTCTACAGAGGCGCTGCGGAAATTCACCA	1330
Db	1231	TCCGGCCTCGTGGCAGGATGTCCACCTCTTTCTACGGAAGGCCCTACGAAATTTACCA	1290
Qy	1331	CCTTCTGATGAAACGGCTAGAAATTTCCCT---CCAGGCTAAACAAAGGGCAAGATTACCCAC	1387
Db	1291	TCCTTCGATGAAACGGCTGGAAATTTCCCTTCAACAGGCTAACAAAGGGCAAGATTACCCAC	1350
Qy	1388	CCCTGACCAAGATTAATGCTATCATAGCTGTGGAGGAGAACATAAATCTTCTCCGATGGA	1447
Db	1351	CCCTGACCAAGATTAATGCTATCACACAGCTGTGGAGGAGAAACACAGCCCTTCTCCGATGGA	1410
Qy	1448	CCCAACATTTGCACAGATAACCCAGCACAGGCGGTGTCCCTCTGCTGCAGACAGGCC---	1504
Db	1411	TCCAAAACATTCACAGAGATAACCCAAACACAGTTTGTTCCTGCTCTGGAGACATTCCTGG	1470
Qy	1505	-----GGAGCCCACTGAGAAAAAACAGAAATTGGTGATTCAGAAAGTTTCAC---	1551
Db	1471	AAACACCCAGGATTCCTCTGAGAAAAAATAGATTAATTAATCTACAGAGTTTCACTTGC	1530
Qy	1552	-----AGAGCCCCAGGGAAGCAGTCTCTTTTGTGAATTAACCGTGGAAAAAGA	1600
Db	1531	TTTGGAAAGAGAGAGCCCTCTGAGGGCTGTCCATCTAGTGAGATACCTATGGAAAAAGGA	1590
Qy	1601	ATGTTGAGAGGACACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCCTTCCTGT	1660
Db	1591	GCCTGGAGAGGGCCGAATAAGTGTAGTTGATTAATCTCATACCTAGAAAGTGACCTTCCCAT	1650
Qy	1661	TTCTACACAGACAGTTTGTAGCAACAACATGATAGATTATATTTTGGAGGCGCCCCAG	1720
Db	1651	TTCTGCCACAGACCACTTGTAGTAACAACTGATAGATTATATTTTGGAGGTGCATCCAG	1710
Qy	1721	TGACTTTGGAAGCCAGCTCTGATTTCTGAAAGTGAGGATTTGGGGCGAGAACTTGAGACGA	1780
Db	1711	TGACCTTGGAAACAAGTTCTGATCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGATGA	1770
Qy	1781	TGGCTTTGATAGCAATGGCTCCCTGCTGTAATCAGACGTGGGAACAGACACTCGGNAAGCCT	1840
Db	1771	TGTTTTTTGATAGTAGTAGTCACTGTGAGACTCAGACTCAGACTTTGAACAAGACCCCTGAAGGCT	1830
Qy	1841	TCACCTTTGGAACCTTTTCCACAGTGTAGATCTCTTACAAACCCCAAACTTTTACAGCCAC	1900
Db	1831	TCACCTTTGGAACTCTTTCTGCGGTGATGATCTCTTTAATCCCAAGAACTTTTACAGCAAC	1890
Qy	1901	GATTCAGACGGCTGCCAGAATTGCCCCAGAGACCCATCAGATTCAGGGAGACATCTGGTGC	1960
Db	1891	AAATTACAGCTGCTGCCAGAAATTTGTTCTGGAAGGCCCTTCTGATTCAGAGAAGGATTTGTC	1950
Qy	1961	TGGCAGCTGTGTTAGGGA---GCTGTACGAGGGGACCCCTTTCGGAGAGACCCCCGACCA	2017
Db	1951	TGGCAAGTCTGATCTAGAGAATTTCTCCAGTCTCTGGAAAGCCCTTCTCTGAGACCCCTGAGCA	2010
Qy	2018	TAGTTCCGGGGAGGAATGACTCGGGAACCGAGTGCAGATCGAGTCAGCAGAGATCTTAAATT	2077
Db	2011	TAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGTCTCAAACT	2070
Qy	2078	GTGGAACTCTTTCTGTCTATCTGAGGACCCCTACAACCTTTTAAATTTTAAAGGCTCCTTT	2137
Db	2071	GTGGAACTCTATCTGTAAATCTGATGACCCCTACAACCCCTTAAATTTTAAAGGCTCCTTT	2130
Qy	2138	TCAACCGTACGGGAAGAAATTTGGAAGGCCGTGAGGACTCAAGGCCCTTCTCTGAGGTAC	2197
Db	2131	TCAAAACATCAGGGGAAAATGAGAAAGGCTGTCTGTGACTCAAAAGACCCCATCTGAGTCCAT	2190
Qy	2198	AGTGGCCTTCTCTGGCCATCATACCTTACTTCTTTGTTAAAGCCACAGCTGTTAGAGGCCA	2257
Db	2191	TGTGGCCATTTCTGAGTGTACACCTTACTTTCTTTTAAAGGTGCAGCTGTTGGGGAGCCA	2250
Qy	2258	AGAAATAAATTTGTCCAGGCTGTGGGCTGGGTGAGGCTCTTGCTGGAGAAAGATACACCCA	2317

Db	2251	AGAAAGTGAATGTCACGACTCGGTACAGCGTGAGCTCTTCTTGAGGAAGACACACACA	23110
Qy	2318	TATCAAGAGAAAAGGTAAACCTTCCTGGGAAGATTTACTGAGTATTATATATAAGTGGTGA	2377
Db	2311	TGTCAAAAGAAAAGGTAAACCTTCCTTGAAGAAGTTACTGAGTATTATATATAAGTGGTGA	2370
Qy	2378	TGAGGATCGCAAAAGGACCATTTGGGAAGAAATTTGCAAGGGATGGATGCAAGTTCCAGAAAACG	2437
Db	2371	TGAGGATCGCAAAAGGACCATTTGGGAAGAAATTTGCAAGGGATGGATGCAAGTTCCAGAAAACG	2430
Qy	2438	AAATTCAGAAAACAGAAAGTTGCCATTGGCTACTGCTTGGCCCTTTTGAGCACACAGAGAAAAGAT	2497
Db	2431	AAATTCAGAAAACAGAAAGTCTATTGGATATTGCTTGACATTTTGAAACACAGAGAAAAGAT	2490
Qy	2498	GTTTAAATAGACT----GAGGATCGAGTCAAAGGACTTTACTGTTGTACAGCAATGTTAAAGA	2553
Db	2491	GTTTAAATAGACTCCAGGGAACATGCTTCAAAGGACTTAATGTTCTCAAGCAATGTTGAGT	2550
Qy	2554	AGTGA-ACAGCCTGCACCCGTGCCACTCTGTCTCTTACTTGAGA-GTTTCCTCTTAAAA	2611
Db	2551	TGGCAGCCTGTAGTCTCTAGCTAGCATACACTACTCTTACCTGAGAGGTGCTCTTTAAAA	2610
Qy	2612	ACAAACACTGGCAGCTGCTCTTGAGACATGTTTTTAAGAAACAACCTTGATCTAGAGATG	2671
Db	2611	ACAAATCTTGGCAGCTGCTCTTTGACATTTTTTTTTTTAGAGGAAATGTAACCTTGGATCT	2670
Qy	2672	CAGTTTCATTATTTTTTGGGTAATGTGTCTCATTAGAA-----ACACCAACT	2717
Db	2671	AGTTTAAATTTTTTTTTTTTGGCAACATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC	2730
Qy	2718	CCGATAATGAAGAAAT-----CTCTTATCTGTAATCCTCTCTTTTCTCTATTTAGTTGGAT	2771
Db	2731	CTGATAATGAAGATGAACCTAGTGTGATTCTAATCCTCCTCTTTTGAATTAGTTGGAT	2790
Qy	2772	GTGGGTTT 2779	
Db	2791	GTGCTTTT 2798	
RESULT 3			
AAH18699 standard; cDNA; 2942 BP.			
XX	AC	AAH18699;	
XX	XX		
DT	26-JUN-2001	(first entry)	
XX	XX		
DE	Human cDNA sequence SEQ ID NO:18964.		
XX	XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
FN	EP1074617-A2.		
XX	XX		
PD	07-FEB-2001.		
XX	XX		
PF	28-JUL-2000; 2000EP-00116126.		
XX	XX		
PR	29-JUL-1999; 99JP-00248036.		
PR	27-AUG-1999; 99JP-00300253.		
PR	11-JAN-2000; 2000JP-00118776.		
PR	02-MAY-2000; 2000JP-00183767.		
PR	09-JUN-2000; 2000JP-00241899.		
XX	XX		
PA	(HELI-) HELIX RES INST.		
XX	XX		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX	XX		
DR	WPI; 2001-318749/34.		
XX	XX		

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 18964; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 2942 BP; 731 A; 741 C; 726 G; 744 T; 0 U; 0 Other;

Query Match 24.2%; Score 1320.8; DB 4; Length 2942;
 Best Local Similarity 72.5%; Pred. No. 0;
 Matches 1962; Conservative 0; Mismatches 662; Indels 84; Gaps 16;

QY 151 CCGTCGCGCGCGCGAGGAGGCTTCTCTATGTGTGAGCGATCTCAACGCGCTAGG 210
 DB 96 CCATCGCGAGGCGGTGGGGGACTCTCTATCCATGGTGTGAGCGCTCGAGCGGACTAG 155
 QY 211 ACGTCTCTTCCCTAGCGGGATGGACTTAACCGGCTCGCCACCGCTTGGCGGGCTC 270
 DB 156 GAACCTCTTCCC-CGCCAGGATGGAGTCCGATCGATCGCGCGCTATTGGCGGGGTGT 214
 QY 271 TGGGCGGTTCGGTGCAGTCTGTTGCGAAGCGCGCTCTCTGCGGCTCTCTGCGCG 330
 DB 215 TCTTCCCTGTGTTCTGCGCGCGCTGCGGATTCGCTGCGCTCTGCGCTTCTGCTGG 274
 QY 331 CGCGGGAATCGGATCGAGTACCACTCCGCTGGCTGGGCAAGCGGAGACTGTGTAGACC 390
 DB 275 CTCGAAGATCGGCTGGAGCAGCAGCCACCGCTGGGCAAGGCGGAGACTCTGTAGGCT 334
 QY 391 TCGGATCCGCTCGGCTGAGCGCGCTGAGCTGTCTCTCTCTCTGCTGAGAGCGCGCC 450
 DB 335 TCCCTCCGAATCCCGTCGACCTCCAGCGCTGAGCGCGCGCGCGCTTACCTGAGAGCTGTC 394
 QY 451 AAG-GAAGAGAGATGGAGACAGGAACGACAGGCGCCCGGAGCGGCTCGGCTCGGCTG 509
 DB 395 AAGAAAAGAGATGGAGCGGAGCAGGCGGATCGCGGAACCGGCTTGGCCCTCGGGCG 454
 QY 510 GGCTCTCGGTTTCGGCTGCCCC---TTCCTTGGCGGATCGCAGCGCTCTCTTCGGAGTTC 566
 DB 455 GGCTTCCGGTTCTGGCCACCCCTTTTCCCTCGGCGATCGCAAGCAGGCTCTTCTTAAGTTC 514
 QY 567 CCGCGGCTTCTCTCGACAAATCCCGGGAATCC-----GCTCTGCC 611
 DB 515 CCGAGCGCTCTTGGCCCGGAAATCTCCGGGAACCCCACTGCTTCTCTCGCCAGCGCC 574
 QY 612 GAGCGTGGAGCCAGTACTGACCAAAATGCTTCTCAGCTCTTGGCTCTGCTCCCTAGC 671
 DB 575 GAGACTCGGTCAGTACTGACGAAACTGCTCTCCAGCTCTTGGCGGCTCTCCCGGA 634
 QY 672 CTATTCCAGAGCTGCTGCTTTTGGAGCCAGCTTTTCCGGGGCGGCTGTATCTCTACAGATGG 731

DB 635 TTGCTTTCAAGAGGTGCTAAATTTGGAGCCAACTTTTGGTGGAAATGTTTCCGACCAGATGG 694
 QY 732 CTAGATTTTGGCCCAAGTTTACAGCGCCCTGAGAGCTTTCGAGAGACGGGAGGAATCTGAC 791
 DB 695 CTAGATTTTGGCTGGAGTCTACAGCGCCCTGAGAGCCCTGAGGAGACGGGAAACCGACC 754
 QY 792 GCTCCCAACGGTGCAGAAAGTCTCTGAGTTTACCTTCGCGGCTGGACTTTCGCGAAGACTCGC 851
 DB 755 GCCCCACACAGCGCAGAAATCTTTGAGTT-CGCTGCACTCGACTCTCTC---AGACCCCTC 810
 QY 852 GTGCTCAGTACTCTTGCATTTGGCTAGAGAGGAGCTCCAGTGCAGCTGCTGCTCTCAGA 911
 DB 811 GGTCAACAGTCCCTTTGATTTGGCTAGAGGAAGGATCCACTGGCAATATCTCGCCCCAGA 870
 QY 912 -CTGGAAGTTTAACTCAAGGCCCAGGAAGAGCTTTAGACTCTGCAAGCGCCCACTTTTCT 970
 DB 871 CCTAAATTTGAGCTTTAAGCCCAAGGAAGTGTCTTTGGACCTTGCAGCAGCAGGCTTTCT 930
 QY 971 CTTGGAGCAGCAGCTGTGGGGAGTGGAGTTGTCTGCCAGTAGCTTTCAAGCTGCTTAGT 1030
 DB 931 CTTTAGAGCAGCAGCTGTGGGGAGTGGAGTGTGTGCCAGTAGCTTTCAATCCCGTCTGTA 990
 QY 1031 CTCCACCGAGAACTTGACTCTTTCATCTCTGGGCTCTGAGGCTTTCAGAGCTTAGGTAA 1090
 DB 991 CTCTAACCGGGAACCTTGGCTCTTTCGCGCTCTGGGCTCTTAAACATTTCAACGCTAGACGA 1050
 QY 1091 TTTTCAAGGTAGTTTCTCTCTCTGAACCTTCTTACCTGAGTACTTCTCCCGAGTTAGG 1150
 DB 1051 TTTTCAAGGTAGTTTCTCTCTCTGAACCTTCTTACCTGAGTACTTCTCCAGGCTAGA 1110
 QY 1151 GCTGCGCTGTCCAGAGCAGCGCTGGAGGTGGCCAGTTTGTGGTTCGGAACACTAACCC 1210
 DB 1111 AGTCAGCTATCAGAAACAGTATGGAATAGCAGAGTAGTGGCTTCCAGACACTAACCC 1170
 QY 1211 AGAGAGTGTATCTTTCTGAAGATGGTGTCTACCTCAGCCGCTTGGGGCAGAGATGTC 1270
 DB 1171 AGAGAGCAGCTGCTGAGAGAGGACCAATTTGTCTATCCCGCGCTGAGTGCAGAACTCAT 1230
 QY 1271 GGCACCGCTGAGAGAGGTGTCGCGCTCTCTCTACAGAGGCTGCGGGAATTCACCA 1330
 DB 1231 TCCGCTCTGTGGCAGGATGTCCACTCTTCTACGGAAGGCTTACAGAAATTCACCA 1290
 QY 1331 CTTCTGTATGAACCGCTAGAAATTCCT---CCAGGCTAACAAAGGCAAGGATTTACCCAC 1387
 DB 1291 TCTTCGATGAACCGCTGGAATTCCTTCAACAGGCTAGCAAGGGCAAGATTTACCCAC 1350
 QY 1388 CCCTGACCAAGATTAATGGCTATCATAGCTTGGAGGAGGAACATAAATCTTCTCCGATGGA 1447
 DB 1351 CCCTGACCAAGATTAATGGCTATCCACAGCTTGGAGGAGGAACACAGGCTTCTCCGATGGA 1410
 QY 1448 CCACAACTTGCACAGATAACCCAGCAGCGGCTGCTCCCTGCTGCGAGCAGAGCC--- 1504
 DB 1411 TCCAAACACTGCAAGATAACCCAAACAGATTTGTTCTGCTGCTGGAGACATTCCTGG 1470
 QY 1505 -----GGAGCCCACTGAGAAAAAACCAGAAATTTGGTGAATTCAGAAAGTTTTCAC--- 1551
 DB 1471 AAACACCCAGGAATCCACTGAAGAAAAAATAGAAATTTAATTAACCTACAGAGGTTCCACTTGC 1530
 QY 1552 -----AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAATTAACCGTGGAAAAAGA 1600
 DB 1531 TTTTGAAGAGAGAGAGCGCTTCTGAGGCTGTCTCATCTAGTGAATACCTATATGGAAGGA 1590
 QY 1601 ATGTGAAGAGGACCACTAAATGCAACTGACCTCTCAGATAGAGGAGAGAGGCTTCTGCT 1660
 DB 1591 GCTGAGAGGCGGAAATAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1650
 QY 1661 TTCTACACAGACCTGTTGTAGCAACAACTGATAGATTAATTTTGGGAGGCGCCCCAG 1720
 DB 1651 TTCTGCGAGCAGCTTGTAGTAACAACTGATAGATTAATTTTGGGAGGTCATCCAG 1710
 QY 1721 TGACTTGAAGCCAGCTCTGATTTCTGAAAGTGAAGATTTGGGGCGAGGAACCTTGAGACGA 1780

Db 1711 TGACCTGGAAACAAAGTTCTGATCCAGAGGCTGAGGATTTGGGATGAGGAAGCTGAGGATGA 1770
Qy 1781 TGGCTTTGATAGCGATGGCTCCCTCTGCTGAATCAGACCTGGAACAGGACTCGGAAGGCT 1840
Db 1771 TGGTTTTGATAGTAGTAGTCACTGTCAGACTCAGACCTTGAAACAGACCTCTGAAGGCT 1830
Qy 1841 TCACCTTTGGAACTCTTTTCCAGAGTATAGATCCCTTACAAACCCCAAACTTTTACAGCCAC 1900
Db 1831 TCACCTTTGGAACTCTTTTCCAGAGTATAGATCCCTTATAATCCCAAGAACTTTTACAGCAAC 1890
Qy 1901 GATTTCAGAGGCTGCCAGAAATGGCCCCAGAGAGCCATCAGATTCAGGGACATCCTGGTC 1960
Db 1891 AATTTCAGACTGCTGCCAGAAATGTTTCTTGAAGAGCCTTCTGATTCAGAGAAGGATTTGTC 1950
Qy 1961 TGGCAGCTGTGTGTAGGGA---GCTGTCCAGAGGAGCCCTTCCGAGAGACCCCGACCA 2017
Db 1951 TGGCAAGTCTGATCTAGAGAATTCTCCAGTCTGGAAGCCTTCTTGAGACCCCTGAGCA 2010
Qy 2018 TAGTTCCGGGAGGAAGATGACTGCGAAACCGAGTGCAGATGAAGCAGAGAATCTTAAAT 2077
Db 2011 TAGTTCTGGGAGGAAGATGACTGCGAAATCTAGTGCAGATGAAGCAGAGAGTCTCAAAC 2070
Qy 2078 GTGGAACTCTTCTGCTCATTCTGAGGACCCCTTACAACCTTTTAAATTTTAAAGGCTCCCTT 2137
Db 2071 GTGGAACTCAATCTGTAATTTCTGATGACCCCTTACAACCCCTTTAAATTTTAAAGGCTCCCTT 2130
Qy 2138 TCAACCGTCAGGGAAGAAATGGAAAGGCGCTCAGGACTCAAAAGGCTCTTCTGAGGTCA 2197
Db 2131 TCAACATCAGGGGAANAATGAGAAGGCTGCTGACTCAAAAGACCCATCTGAGTCCAT 2190
Qy 2198 AGTGCCCTTCTGTGCCATCATACCTTAATCTTTGTAAGGCCAGCCAGCTGTGTAGAGGCCA 2257
Db 2191 TGTGGCCATTTCTGAGTGTCAACCTTACTTTCTTGTAAAGGTGCAGCTGTTGGGGAGCCA 2250
Qy 2258 AGAGATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTCTTCTGAGAGAAATACACCA 2317
Db 2251 AGAAAGTGAATGTCAGACTCGGTACAGCGTGACGTTCTTTCTGAGGAAGACACACA 2310
Qy 2318 TATCAAGAGAAAAAGGTAACCTTCTCGGAAGAGTTACTGAGTATTTATATAAGTGSTGA 2377
Db 2311 TGTCAAAAGAAAAAGGTAACCTTCTTGAAGAGTTACTGAGTATTTATATAAGTGSTGA 2370
Qy 2378 TGAGATCGCAAGAACCATGGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2437
Db 2371 TGAGATCGCAAGAACCATGGGAAGAAATTTGCAAGGATGGATGCGAGTTCCAGAAACG 2430
Qy 2438 AATTCAAGAAACAGAAAGTTGCAATTTGGCTACTGCTTGGCTTTGAGCAGCAGAGAAAAAT 2497
Db 2431 AATTCAAGAAACAGAAAGTCTATTTGGATATTTGCTTGACATTTGAACACAGAGAAAGAT 2490
Qy 2498 GTTTAATAGACT---GAGGATCGAGTCAAAAGGACTTACTGTTGTACAGCAATGTTAAGA 2553
Db 2491 GTTTAATAGACTCCAGGGAACATGCTTCAAAAGGACTTAATGTTCTCAAGCAATGTTGAT 2550
Qy 2554 AGTGA-ACAGCTGCAACCCGTGCCACTCTGCTCTTACTTGTAGA-GTTTCCCTTAAAA 2611
Db 2551 TGGCAGCCTGTAGTCTAGTAGCATACACTACCTCTTACCTGAGAGTGCTTTTAAAA 2610
Qy 2612 ACAACACTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAATTTGATCTAGAGATG 2671
Db 2611 ACAAACTTGGCAGCTGTCTTGGACATGTTTTTAAAGAAACAATTTTATAGAGAAATGAACTTGGATCT 2670
Qy 2672 CAGTTTGAATATTTTGGGTAATGTGCTCATTTAGAA-----ACACCAACT 2717
Db 2671 AGTTTAAATTTTTTTTTTGGCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC 2730
Qy 2718 CGATAATGAAGAAAT-----CTCTTATCTGTAACTCTCTCTTTTCTATTTAGTTGGAT 2771
Db 2731 CTGATAATGAAGGATGAAC TAGTGTGATTTCTTAATCTCTCCCTTTTGTGATTTAGTTGGAT 2790
Qy 2772 GTGGGTTT 2779
Db 2791 GTGCTTTT 2798

RESULT 4

ID AAC98158 standard; cDNA; 2324 BP.
XX AAC98158;
XX AC AAC98158;
DT 09-MAR-2001 (first entry)
XX Human colon cancer antigen nucleotide sequence SEQ ID NO:168.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX identification; cytostatic; cardioactive; neuroprotective; vulnerary;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;
XX infectious disease; cardiovascular disorder; ss.
XX Homo sapiens.
XX W0200055351-A1.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US005883.
XX 12-MAR-1999; 99US-0124270P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2000-587534/55.
XX P-PSDB; AAB53401.
XX Colon cancer associated gene sequences, referred to as colon cancer
XX antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer.
XX Claim 1; Page 595; 2104pp; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnerary, nephrotropic, antiinfective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies to the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins may
XX also be used to prevent diseases such as neural disorders, immune system
XX disorders, muscular disorders, reproductive disorders, gastrointestinal
XX disorders, wounds, renal disorders, infectious diseases, and
XX cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
XX sequences used in the exemplification of the present invention
XX Sequence 2324 BP; 615 A; 566 C; 562 G; 574 T; 0 U; 7 Other;

Query Match 21.7%; Score 1189.2; DB 3; Length 2324;
Best Local Similarity 75.3%; Pred. No. 1.2e-295;
Matches 1642; Conservative 4; Mismatches 477; Indels 59; Gaps 11;

Qy 519 TTCCGGCTGCCTTCCTTCGGCGATCGACGCTGCTTCGGAGTTCCCGCGCTTCC 578
Db 43 TTTCAACCCCTGTGTAGTCGGCGATCGAAGAGGCTCTTCTAAGTTCGACGCTCTT 102
Qy 579 TCTCGACAAAATCCCGGAACTCCGCTCT-----GCCCGAGCGTCGAC 623
Db 103 GGCCCGGAAAACCTCGGGAACCCACACTGCTTCTCTGCCCGCGAGACTCGGGTC 162

Qy 624 AGGTACTGGACCAAAATGCTTTCTCAGCTCCTTGCCTGCTCCTAGCCTATTCAGAAAG 683
 Db 163 AGTTACTGGACCAAAATGCTCCTCCAGCTCCTTGGCGCTCCCGGATTCCTCAGAAAG 222
 Qy 684 CTGCTGCTTTGGAGCCAGCTTTCCGGGGCTGATTCCTACCAAGATGGCTAGATTTGGCC 743
 Db 223 GTGCTAAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCGGACCAATGGCTAGATTTGCT 282
 Qy 744 GCAAGTTACAGCGCCCTGAGAGCTTCGAGAGGACGGGAGGAATTCAGCGCTCCACCGGTG 803
 Db 283 GGAGTCTACAGCGCCCTGAGAGCCCTGAAGGGACGGGAGAAACAGCGCCGCCACAGCG 342
 Qy 804 CAGAAGTCTCTGAGTTACATGCGCGCTGAGACTCTTCGGGAAGACTCGCGTGTGCTAGTACT 863
 Db 343 CAGAAATCTTTGAGTT-CGCTGCACTCGACTCCTC---AGACCCCTCGGTCAACAGTCC 398
 Qy 864 CTTGCATTTGGCTAGAGAGGAGCTCCAGTGGCAGCTGCTGCTCCTCAG-CTGGAAGTTAA 922
 Db 399 CTTGATTTGGCTAGAGAGGGGATCCACTGGCAATACTTCGCCCCAGACCTTAAATTTGGA 458
 Qy 923 ACTCAAGGCCCAGGAAAGAGCTTTAGACTCTGCAAGCGCCCACTTTCTCTGAGAGCAGCA 982
 Db 459 GCTTAAGGCCAAGGAAGTGTCTTGGACCTGCAAGCAGAGCTTTCTCTTAGAGCAGCA 518
 Qy 983 GCTGTGGGAGTGGAGTTGCTGCCAGTAGAGCTTCAAGCTTCAAGCTGGTCTAGTCTCCCAACGAGA 1042
 Db 519 GCTGTGGGAGTGGAGCTGTTGCCAGTAGAGCTTCAATCCCGTCTGTACTCTAACCCGGA 578
 Qy 1043 ACTTGACTCTTCATCTCTGGGCTCTGAGGGTTGAGGCTTCAAGCTTAGGTAATTCAGGTAGT 1102
 Db 579 ACTTGGCTCTTGGCCCTTGGGCTTCTAAACATTCAGCGCATAGCAATTTCAGTGTGGT 638
 Qy 1103 TTCTATCTCTGTAACCCCTTCTACCTGGACTACTTCTCCCACTAGGGCTGCGCTGTCA 1162
 Db 639 ATCTATTTGCTGAACCCCTTCTACTGGACTGCTTCTCTAGGCTAGAGTCACTATCA 698
 Qy 1163 GAGCAGCGCTGGAGGTGGCAGTTGTTGGGTTTCGGAACACTAAACCCAGAGAGCTGCTA 1222
 Db 699 GAAACAGTGATGGAATAGCGAGGTAGTGGCTTCCAGACACTAAACCCAGAGAGCAGCTG 758
 Qy 1223 TCTTTCTGAGATGTTGTACCTCAGCCGCTTGGGGCAGAGATGTCGGCAACCGCTG 1282
 Db 759 CCTGAGAGAGACCATTTGTATCCCGCAGCGCTGARTGCAGAACTCAATTCGGCTCGTG 818
 Qy 1283 GAGAAGGTGTCGGCTCTCTACAGAGGCGCTCCGGAATCCACCACTTCCTGATGAA 1342
 Db 819 GCAGGATGTCACCTCTTCTACGGAAGGCTTACAGAAATTCACCATCTTCGATGAA 878
 Qy 1343 ACGGCTAGAAATTCCT---CGAGGTAAACAAAGGGCAAGAGTTACCCACCCCTGACCAAGA 1399
 Db 879 ACGGCTGGAATTCCTTCAACAGGCTAGCAAGGGGCAAGATWTACCCACCCCTGACCAAGA 938
 Qy 1400 TAATGGCTATCATAGCCTGGAGGAGGACATACTTCTCGGATGACCCACACATTTG 1459
 Db 939 TAATGGCTTACCAAGCCTGGAGGAGGAAACAGAGCTTCTTCGGATGATCCAAACACTG 998
 Qy 1460 CACAGATAACCCAGCACAGGCGGTGCTCCCTGCTGCAGACAGGCG---GGA 1507
 Db 999 CAGAGATAACCAACACAGTTTGTCTGCTGCTGGAGACATTCCTGGAAACACCCAGGA 1058
 Qy 1508 GCCCACTGAGAAAAACAGAAATTTGGTATTCAGAAAGTTTCAC-----A 1552
 Db 1059 ATCCACTGAGAAAAATAGAAATTAATACTACAGAGGTTCCACTTGTCTTGGGAAGA 1118
 Qy 1553 GAGCCCCGAGGAGCAGTCTGTTTGTGATTAACCGTGGAAAGAAATGTAAGAGGA 1612
 Db 1119 GAGCCCTTCTGAGGCTGTCCATCTAGTGAATACCTATGGAAGAGGAGCCTGAGAGGG 1178
 Qy 1613 CCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCTTCTGTTTCTACAGACC 1672
 Db 1179 CCGAATAAGTGTAGTTGATTACTCATACCTAGAGAGGTGACCTTCCCATTTCTGCAGACC 1238
 Qy 1673 AGTTGTAGCAAACTGATAGATTAATTTTGGAGGCGCCGCCAGTGTGGAAGC 1732

Db 1239 AGCTTGTAGTAAACAACTGATAGATTAATTTTGGAGGTGTCATCCAGTACCTGGAAAC 1298
 Qy 1733 CAGCTCTGATTTCTGAAAGTGAGGATTTGGGCGGAGGAACCTTGAGGACGATGGCTTTGATAG 1792
 Db 1299 AAGTTCTGATCCAGAGAGGTGAGGATTTGGGATGAGGAGCTGAGGATGATGTTTGTATAG 1358
 Qy 1793 CGATGGCTCCCTGCTGTGAATCAGACGTGGAACAGGACTCTGGAAGGCTTCCCTTTGGAA 1852
 Db 1359 TGATAGCTCACTCTGACAGCTCAGACCTTGAACAGACCCCTGGAAGGCTTCCACCTTTGGAA 1418
 Qy 1853 CTCTTTCCACAGTGTAGTCCCTTACAAACCCCAAACTTTTACAGCCACGATTCAGAGCGC 1912
 Db 1419 CTCTTTCTGAGTGTAGTCCCTTATATCCCAAGAACTTTTACAGCAACAAATTCAGACTGC 1478
 Qy 1913 TGCAGAAATTTGCCCCACAGACCCATCAGATTCAGGGAACATCCTGTCTGGCAGCTGTGG 1972
 Db 1479 TGCAGAAATTTGCTTCCAGAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCAAGTCTGA 1538
 Qy 1973 TGTAGGGA---GCTGTGAGAGGGAACCCCTTCCGAGAGACCCCGACCATAGTTCGCGGGA 2029
 Db 1539 TCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGAGCATAGTTCTGCGGA 1598
 Qy 2030 GGAAGATGACTGGGAACCCAGTGCAGATGAGCAGAGAACTCTTAAATTTGTGGAACTCTTT 2089
 Db 1599 GGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTATT 1658
 Qy 2090 CTGTCTTCTGAGGACCCCTTACAAACCTTTTAAATTTTAAAGCTCTCTTTCAACCGTCAGG 2149
 Db 1659 CTGTAAATCTGATGACCCCTTACAAACCTTTTAAATTTTAAAGCTCTCTTTCAACATCAGG 1718
 Qy 2150 GAAGAAATTTGAAAGCGCTCAGAGCTCAAAAGGCTCTTCTGAGGTGACAGTGGCTTCTC 2209
 Db 1719 GGAATATGAAAGGCTGCTGCTGACTCAAAAGACCCATCTGAGTCCATTTGGGCCATTTTC 1778
 Qy 2210 TGCCCATCATACCTTACTTTCTTGAAGGCCAGCTGTTAGAGAGCCCAAGAAATATTG 2269
 Db 1779 TGAGTGTCAACCTTACTTTCTTGAAGGTGACAGTGTGGGAGGCCAAGAAAGTGAATG 1838
 Qy 2270 TCAGAGCTGTGGGCTGGGTGAGGCTCTTGTGAGAGAAAGATACACCCATATCAAGAGAAA 2329
 Db 1839 TCAGAGCTGGTACAGGTTGACGTTCTTCTGAGGAGAGACACACATGTCAAAGAAA 1898
 Qy 2330 AAAGGTAACCTTCTCGAAGAACTTACTGATGATTAATAAGTGGTGTAGAGATCGCAA 2389
 Db 1899 AAAGGTAACCTTCTTGAAGAGTTACTGAGTATTAATAAGTGGTGTAGAGATCGCAA 1958
 Qy 2390 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGAGTTCAGAAACGAAATTCAGAAAC 2449
 Db 1959 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGAGTTCAGAAACGAAATTCAGAAAC 2018
 Qy 2450 AGAAGTTGCCATTTGGCTTACTGCTTGGCTTTGAGCACAGAGAAAATGTTTAAATAGACT 2509
 Db 2019 AGAAGATGCTATTTGGATATATGCTTGATATTTGAAACACAGAGAAAGAAATGTTTAAATAGACT 2078
 Qy 2510 ----GAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTAAGAAAGTGA-ACAGCC 2564
 Db 2079 CCAGGGAACATGCTTCAAAGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTA 2138
 Qy 2565 TGCAACCCGTCGCCACTCTGCTCTTACTTTGAGA-GTTTCCCTTAAAAACAAACACTGGC 2623
 Db 2139 GTCTTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTCTTTTAAAAACAAATCTTTGGC 2198
 Qy 2624 AGCTGTCTTTGGACATGTTTTT 2645
 Db 2199 AGCTGTCTTTGACATTTTTTT 2220

RESULT 5
 AAH99389
 ID AAH99389 standard; cdna; 883 BP.
 XX
 AC AAH99389;

XX 16-OCT-2001 (first entry)
 XX Human protein encoding cDNA sequence SEQ ID NO:224.
 XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; anaemia;
 KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 XX Homo sapiens.
 XX WO200153455-A2.
 XX 26-JUL-2001.
 XX 22-DEC-2000; 2000WO-US035017.
 XX 23-DEC-1999; 98US-00471275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-00552317.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-457603/49.
 DR P-PSDB; AAM25448.
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
 XX Claim 1; Page 398; 1217pp; English.
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and cells
 CC they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders
 XX
 SQ Sequence 883 BP; 263 A; 173 C; 218 G; 229 T; 0 U; 0 Other;
 Query Match 9.9%; Score 542.6; DB 4; Length 883;
 Best Local Similarity 80.5%; Pred No. 5,8e-129;
 Matches 688; Conservative 0; Mismatches 154; Indels 13; Gaps 4;

QY 1719 AGTGACTTGGAGCCAGCTCTGATTTCTGAATCTGAAGTAGAGGATTGGGGCGAGGAACTCTGAGGAC 1778
 DB 1 AGTGACCTGGAAAACAAGTTCTGATCTCAGAAAGGTGAGGATTTGGGATGAGGAAGCTGAGGAT 60
 QY 1779 GATGGCTTTGATAGCGATGGCTCCCTGCTCTGTAATCAGACCTGGACAGGACTCGGAAGGC 1838
 DB 61 GATGGTTTGTATAGTAGTATGCTCAGCTGTAGAGCTCAGACCTTGAACCAAGACCCCTGGAAGGG 120
 QY 1839 CTTCAACCTTTGGAACTCTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTTACAGCC 1898
 DB 121 CTTCAACCTTTGGAACTCTTTCTGCAGTGTAGATCTTTATATATCCCAAGACTTTTACAGCA 180
 QY 1899 ACGATTACAGCGGCTGCCAATTTGCCCCAGAGACCCATCAGATTCAGGAGACATCTCTGG 1958
 DB 181 ACAATTACAGACTGCTGCCAGAAATTTGTTCTCTGAAGAGCCTTCTGATTTCAGAGAAGGATTTG 240
 QY 1959 TCTGGCAGCTGTGGTGTAGGGA---GCTGTTCAGAGAGGACCCCTTCCGGAGACCCCGAC 2015
 DB 241 TCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTCGAAGCCTTCTCTGAGACCCCTGAG 300
 QY 2016 CATAGTTCCGGGAGGAGGATGACTGGGAAACCCGAGTGCAGATGAAGCAGAGAAATCTTTAAA 2075
 DB 301 CATAGTTCTGGGAGGAGGATGACTGGGAAATCTAGTGCAGATGAAGCAGAGAGTCTTCAA 360
 QY 2076 ---TTGTGGAACCTTTCTGTCTCA---TTCTGAGGACCCCTTACAACTTTTAAATTTTAAAG 2129
 DB 361 ACTGTGGGAACTTCAATCTGTAAATCTGGATGGACCCCTTACAACTTTTAAATTTTAAAG 420
 QY 2130 GCTCCTTTTCAACCGTCAGGAGGAGAAATTTGGAAGGCGTCAGGACTCAAGAGGCTCTTCT 2189
 DB 421 GCTCCTTTTCAACATCAGGAGGAGAAATGGAAGAGGCTGTCTGACTCAAGAGACCCCTTCT 480
 QY 2190 GAGTCAAGTGGCCTTCTCTGGCCATCATACCTTACTTTCTTGAAGGCCAGCTGTTA 2249
 DB 481 GAGTCCATTTGGGCCATTTCTGAGTGTACACCTTACTTTCTTGAAGGTCAGCTGTG 540
 QY 2250 GAGAGCCAAAGAAATATTTGTCAGGCTGTGGGCTGTGGGCTGTGGGCTCTTCTGTGAGAAAGA 2309
 DB 541 GGGAGCCAAAGAAATGTAATGTCCAGACTCGGTACAGCGTGACGTTCTTCTTGGAGGGAAGA 600
 QY 2310 TACACCATATCAAGAGAAAAGGTAACCTTCTCTGGAAGAGTTACTGTAGTATTATATA 2369
 DB 601 CACACATGTCAAAAGAAAAGGTAACCTTCTTGAAGAGTTACTGTAGTATTATATA 660
 QY 2370 AGTGGTGTATGAGGATCGCAAGGACCATGGGAAGAAATTTCAAGGGATGGATGAGGTTT 2429
 DB 661 AGTGGTGTATGAGGATCGCAAGGACCATGGGAAGAAATTTCAAGGGATGGATGAGGTTT 720
 QY 2430 CAGAAACGAATTCAGAAAACAGAAAGTTGCCATTTGGCTTACTGTCTTGGCTTTGAGCAGAGA 2489
 DB 721 CAGAAACGAATTCAGAAAACAGAAAGTGTATTTGGATATTTGCTTGAATTTGAACACAGA 780
 QY 2490 GAAAAATGTTTAAATAGACT---GAGGATCGAGTCAAGAGACTTACTGTGTACAGCAA 2545
 DB 781 GAAAGAAATGTTTAAATAGACTCCAGGGAACATGCTTCAAGAGACTTAAATGTTTCTCAAGCAA 840
 QY 2546 TGTTAAGAACTGAAC 2560
 DB 841 TGTGAGTTGGCAGC 855
 RESULT 6
 AAH06623
 ID AAH06623 standard; cDNA; 772 BP.
 XX
 AC AAH06623;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:3458.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX

PT tissues or cells.
 PS Claim 1; SEQ ID NO 1296; 239pp; English.
 XX The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilise a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterised by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent

XX Sequence 540 BP; 162 A; 119 C; 85 G; 174 T; 0 U; 0 Other;
 SQ

Query Match 3.3%; Score 181.6; DB 6; Length 540;
 Best Local Similarity 75.8%; Pred. No. 6.8e-36;
 Matches 388; Conservative 0; Mismatches 69; Indels 55; Gaps 11;

QY 4960 ATATTATAAAATTTTCAGATTACCTTCACTATCAAAATGAGTAATGAC----- 5011
 DB 499 ATGTTGAAATTTTCCAGATTATACCTCACTATCAAAATGAAATGACTGGTTAAA 440

QY 5012 ----TCATGCTGCAGACATGTCCTGATGTGGCAAGACAGAGGATCT-----TTG 5060
 DB 439 ATAATCAGCCCTGCAGAGATGTCAAATGGTGGCAAGACAGAGGGTCTAACTGAAGG 380

QY 5061 ACTGAAGGAGAAAACTGTCATTGTCATCCAGCCCCAGGAAAGAACACCTCCAAAGGCA 5120
 DB 379 AGAAAAACAAAATTAATCTGTCATTGTCAT-CGAAATCCCAAGGAAAGAACACCTCCGAGGCA 321

QY 5121 GCGAGCGAGGCGAGCGAGCATGGTGGTCTAGTTGAATACATTC-AAGTCTTGCAATG 5179
 DB 320 -----AGGCAGCGCATGGCGGTCTAGCTGAACACTTACACTCAGTCTTGCAATG 272

QY 5180 GTGCTTT--AGATCTGTAGCATGTGAGGCTCTGTACAGTGGGG-CCACACTTCTGA 5236
 DB 271 GTGCTTTAGAGATCTCTGTAGCATGTGAGGCTCTGTACAGTGGGGACCCACACTTCCGA 212

QY 5237 GGGCTGGAATGTGGCAACCTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTT 5296
 DB 211 GTGCTGGAATGTGGCAACCC-TTATCTAACTTGAATCAAAACCGTCAAGATTTATTTTT 153

QY 5297 TATAATTTAAGAA-----AGAGTTGGGGAATGACATTTTTTGAAGTTGGCTTTTTCAGCTCA 5352
 DB 152 TATAATTTAAGAAAAAGTAGACACGGAATGAGAAATTTTTTTTTTTTGAATGAGATGTT 93

QY 5353 G-----TCATTTTACGTGTAAAC--GTGAGATTTTGTAGCTCAGATTTATATT 5397
 DB 92 GAGTTGGTTTCAGATCATTTTACATGTAACTAGTGTAGATTTTATAGCGGAGATTATATT 33

QY 5398 TGTATATAATTTAATACTAATCTGTAATTTGT 5429
 DB 32 TGTATATAATTTAATACTAATCTGTAATTTGT 1

RESULT 8
 ADB57750/C
 ID ADB57750 standard; DNA; 540 BP.
 XX
 AC ADB57750;
 XX
 DT 04-DEC-2003 (first entry)
 DE Toxicity-related gene, SEQ ID 2776.
 XX
 KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
 drug screening; toxicity assay; ds.
 XX
 OS Unidentified.
 XX
 PN WO2003064624-A2.
 XX
 PD 07-AUG-2003.
 XX
 PF 31-JAN-2003; 2003WO-US003194.
 XX
 PR 31-JAN-2002; 2002US-00060087.
 PR 15-MAR-2002; 2002US-0364045P.
 PR 15-MAR-2002; 2002US-0364055P.
 PR 30-DEC-2002; 2002US-0436643P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 DR WPI; 2003-689530/65.
 XX

Predicting a toxic effect of a compound, useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays, comprises preparing gene expression profile of tissue or cells exposed to the compound.

Claim 1; SEQ ID NO 2776; 1156pp; English.

The present invention relates to a method for predicting a toxic effect of a compound. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the gene expression profile to a database comprising SEQ ID 1-4925, where differential expression of the gene indicates at least one toxic effect. The method is useful for predicting at least one toxic effect of a compound, predicting hepatotoxicity or the progression of a toxic effect of a compound, identifying an agent that modulates the onset or progression of a toxic response, predicting the cellular pathways that a compound modulates in a cell, and identifying an agent that modulates at least one activity of a protein. The method and compositions of the present invention using a database of genes having liver toxin-induced differential expression, are useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 540 BP; 162 A; 119 C; 85 G; 174 T; 0 U; 0 Other;

Query Match 3.3%; Score 181.6; DB 10; Length 540;
 Best Local Similarity 75.8%; Pred. No. 6.8e-36;
 Matches 388; Conservative 0; Mismatches 69; Indels 55; Gaps 11;

QY 4960 ATATTATAAAATTTTCCAGATTACACTTCACTATCAAAATGAGTAATGAC----- 5011
 DB 499 ATGTTGAAATTTTCCAGATTATACCTCACTATCAAAATGAAATGACTGGTTAAA 440

QY 5012 ----TCATGCTGCAGACATGTCCTGATGTGGCAAGACAGAGGATCT-----TTG 5060

Db 439 ATATCAAGCTGAGAGTGTCAATGGTGGCAAGACAGAGAGGTCTAAACTGAAGG 380
 Qy 5061 ACTAAGGAGAAATGTCATTCTCATCCAGCCGCCAGGAAAGAACACTCCCAAGGCA 5120
 Db 379 AGAAAACAAATTAATGTCATTCTCATCCAGGAAAGAACACTCCGAGGCA 321
 Qy 5121 GGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5179
 Db 320 -----AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 272
 Qy 5180 GTGCTTTT--AGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGG--GCCACATTTCTGA 5236
 Db 271 GTGCTTTTGTAGAGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGGAGCCACACATTTCCGA 212
 Qy 5237 GGGCTGAAATGTGCAACCCCTTTTAACTTGAATCAAAACCGTCAAAATTTTATTTTT 5296
 Db 211 GTGCTGAAATGTGCAACCC--TTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTT 153
 Qy 5297 TATAATTTAAGAA-----AGAGTTGGGAATGACATTTTTTTGAGTTGGCCCTTTTCAGCTCA 5352
 Db 152 TATAATTTAAGAAAGTAGAGACGGAATGAGAAATTTTTTTTTTTTGGAAATGAGATGTTT 93
 Qy 5353 G-----TCATTTTACGTGAAC--GTGGAGATTTGATAGCTCAGATTATTT 5397
 Db 92 GAGTTGGTTTTCAGATCACTTTTACATGTAACTAGTGTAGATTTTATGCGGAGATTATTT 33
 Qy 5398 TGTATATAATTTAATTAATCTTAATCTGTAATTTGT 5429
 Db 32 TGTATATAATTTAATTAATCTGTAATTTGT 1

RESULT 9
 ID ABX52167 standard; cDNA; 399 BP.
 AC ABX52167;
 DT 25-FEB-2003 (first entry)
 DE Bovine EST associated with lactation/muscle/fat deposition #2096.
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 OS Bos Taurus.
 PN US2002137160-A1.
 PD 26-SEP-2002.
 PP 26-OCT-2001; 2001US-00983965.
 PR 17-DEC-1998; 98US-0113678P.
 PR 15-DEC-1999; 99US-00465231.
 PA (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 DR WPT; 2003-102386/09.
 XX Purified nucleic acid molecules, useful for genome mapping, gene
 PT identification and analysis, cattle breeding or preparation of constructs
 PT for cattle gene expression and genetically improved cattle.
 PS Claim 2; SEQ ID NO 2096; 38pp; English.
 CC The invention relates to a purified nucleic acid molecule associated with

CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
 CC nucleic acid molecule comprising any of 5912 nucleotide sequences,
 CC appearing as ABX50072-ABX5983, or complements of them. Also included are
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 5912 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridization between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the 5912
 CC bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present
 CC sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137160
 XX
 SQ Sequence 399 BP; 133 A; 62 C; 82 G; 122 T; 0 U; 0 Other;
 Query Match 3.1%; Score 170.8; DB 8; Length 399;
 Best Local Similarity 79.1%; Pred. No. 3.5e-33;
 Matches 231; Conservative 0; Mismatches 52; Indels 9; Gaps 2;
 Qy 2330 AAAGGTAACCTTCTCGAAGAAGTTACTGAGTATTATATAAGTGGTGAAGATCGCAA 2389
 Db 102 ATAGGTAACCTTCTCGAAGAAGTTACTGAGTATTATATAAGTGGTGAAGATCGCAA 161
 Qy 2390 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGATGATGATGATGATGATGATG 2449
 Db 162 AGGACCATGGGAAGAAATTTGCAAGGATGATGATGATGATGATGATGATGATGATG 221
 Qy 2450 AGAAGTTGCCATTTGGCTACTGCTTGGCTTTGAGCAGACAGAGAGAGAGAGAGAG 2502
 Db 222 AGAAGAGCTATTTGGATGATGCTTTGATGATGATGATGATGATGATGATGATGATG 281
 Qy 2503 ATAGACTGAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTAAGA--AGTGAAC 2560
 Db 282 CTTCAAGCAACATGTTTCAAGGACTTAATGTTTTCGAGCAATGTTAAGATGATTCGAC 341
 Qy 2561 AGCCTGCAACCCGTCGCTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2612
 Db 342 AGCCTCTGGCCCTAGCATACACTACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 393
 RESULT 10
 AA196694/c
 ID AA196694 standard; cDNA; 689 BP.
 AC AA196694;
 DT 13-NOV-2001 (first entry)
 DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2769.
 DE Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
 KW Homo sapiens.
 OS
 XX WO200166719-A1.
 XX
 PD 13-SEP-2001.
 PP 02-MAR-2001; 2001WO-JP001629.
 CC

```

PR 07-MAR-2000; 2000JP-00159195.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
DR WPI; 2001-565584/63.
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT malignancy and susceptibility indicator or tumor marker for anti-cancer
PT agents.
XX
XX Claim 1; Page 2028; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes
XX
SQ Sequence 689 BP; 221 A; 114 C; 93 G; 251 T; 0 U; 10 Other;
Query Match 2.1%; Score 113.6; DB 4; Length 689;
Best Local Similarity 72.3%; Pred. No. 2.8e-18;
Matches 191; Conservative 0; Mismatches 64; Indels 9; Gaps 3;
QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271
DB 293 ACAGGTAAGAGTTCCATTTCTGAGTGATGAATGTAACTCTTTCATCTTAACTTGA 234
QY 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGCTTGGGGAATGAC 5325
DB 233 ATCAAAACCTATCAGATTTTATTTTGTATATAATTTAAGGAAGTAAAGTTAGGGGACTAGA 174
QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACTGGAGATTGTATAG 5384
DB 173 AGACTCTAAATTTGGCTTCTACAGATCAATAATTTAAATGTAACTAGTTGGGATTTATAG 114
QY 5385 CTCAGATTATATTTGTATATATAATTAATACTAACTCTGTAATGTGTAATAATAATATTGTC 5444
DB 113 TTAATAATATATTTCTGTATATAACATAACTAACTCTGTAATGTGTAATAATAATATTGTC 54
QY 5445 AATTATTAAAAAAGGAAAAA 5468
DB 53 AATTATTAAAAAAGGAAAAA 30
RESULT 11
ACH49151
ID ACH49151 standard; cDNA; 421 BP.
XX
XX ACH49151;
XX
XX 13-OCT-2003 (first entry)
XX
XX Human leukocyte cDNA #745.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX

```

```

PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 36363; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 421 BP; 158 A; 50 C; 69 G; 144 T; 0 U; 0 Other;
Query Match 2.0%; Score 109; DB 9; Length 421;
Best Local Similarity 71.6%; Pred. No. 3.2e-17;
Matches 187; Conservative 0; Mismatches 65; Indels 9; Gaps 3;
QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271
DB 132 ACAGGTAAGAGTTCCATTTCTGAGTGATGAATGTAACTCTTTCATCTTAACTTGA 191
QY 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGCTTGGGGAATGAC 5325
DB 192 ATCAAAACCTATCAGATTTTATTTTGTATATAATTTAAGGAAGTAAAGTTAGGGGACTAGA 251
QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACTGGAGATTGTATAG 5384
DB 252 AGACTCTAAATTTGGCTTCTACAGATCAATAATTTAAATGTAACTAGTTGGGATTTATAG 311
QY 5385 CTCAGATTATATTTGTATATAATAATTAATACTAACTCTGTAATGTGTAATAATAATATTGTC 5444
DB 312 ATAAATTTATATTTGCGTATATAACATAACTAACTCTGTAATGTGTAATAATAATATTGTC 371
QY 5445 AATTATTAAAAAAGGAAAAA 5465
DB 372 AATTATTAAAAAAGGAAAAA 392
RESULT 12
AAH11666/c
ID AAH11666 standard; cDNA; 543 BP.
XX
XX AAH11666;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (3'-primer) SEQ ID NO:8501.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX

```

XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX XX
 XX PD 07-FEB-2001.
 XX XX
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX XX
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00300253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00183767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX XX
 XX PA (HELI-) HELIX RES INST.
 XX XX
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX XX
 XX DR WPI; 2001-318749/34.
 XX XX
 XX XX
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX PT
 XX XX
 XX PS Claim 3; SEQ ID NO 8501; 2537pp + Sequence Listing; English.
 XX XX
 XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX XX
 XX SQ Sequence 543 BP; 189 A; 104 C; 97 G; 149 T; 0 U; 4 Other;
 XX XX
 XX Query Match 2.0%; Score 109; DB 4; Length 543;
 XX Best Local Similarity 63.7%; Pred. No. 3.8e-17;
 XX Matches 254; Conservative 0; Mismatches 119; Indels 26; Gaps 5;
 XX XX
 XX QY 2407 TTGCAGGGATGGATCGAGTTCAGAACCAATTCAGAACAGAGATGCCATTCGCT 2466
 XX DB 543 TTTTCAAGGATGAATCCAGTTCNAGAAAGATTCAGAACAGAGATGCTATTGGAT 484
 XX QY 2467 ACTGCTTGGCCCTTTGAGCACAGAGAAAATGTTTATAGACT----GAGGATCGAGTCA 2522
 XX DB 483 ATGGCTTGACATTTGACCCAGAGAGAAAGATGTTTATAGACTCCAGGACATGCTTCA 424
 XX QY 2523 AAGGACTTACCTGTTGACAGCAATGTTAAGAAAGTGA-ACAGCTGCAACCCGTCGCCACT 2581
 XX DB 423 AAGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACA 364
 XX QY 2582 CTGTCTCTTACTTGAGA-GTTTCCCTTAAACAAACAAACACTGGCAGCTGCTCTTGACATG 2640
 XX DB 363 CTACCTCTTACCTGAGAGGTGCTCTTTTAAAAAACAATCTTGGCAGCTGCTCTTGACATT 304

QY 2641 TTTTAAAGAAACAACACTTGATCTAGAGATGCGAGTTTGATTTATTTTGGTAATGTGTCT 2700
 DB 303 TTTTATTTAGAGGAAATGTAACCTGGATCTAGTTTAAATTTTATTTTTCACAATATCC 244
 QY 2701 CATTAGAA-----ACACCAACTCCGATAATGAGAAAT-----CTCTTATC 2740
 DB 243 CACTCAGAAACATTCAGGTTTGAAGCCGCCCTGATAATGAAGATGAAGTAAGTAGTGATT 184
 QY 2741 TGTAAATCCTCTCTTTTCTTATTTAGTTGGATGTTGGTTT 2779
 DB 183 TCTAATCCTCCCTTTTGTGTTTATTTAGTTGGATGTTGCTTTT 145
 RESULT 13
 AAF21761
 ID AAF21761 standard; DNA; 1028 BP.
 XX
 AC AAF21761;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 148.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive; neotropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; antitumor; antitumor; antitumor; antibacterial; antifungal; antiparasitic; cardiac; immune disorder; Addison's disease; allergy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease; ds.
 KW
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005881.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR P-PSDB; AAB58858.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the polypeptides encoded by these genes, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases.
 PT
 PS Claim 1; Page 592; 1299pp; English.
 XX
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human proteins AAB58711 - AAB59128. The DNA and protein sequences are associated with breast and ovarian cancer. Included in the invention are sequences AAF22032 - AAF22040 and AAB59129 which are used in the invention. The breast and ovarian cancer associated DNA, protein, agonist or antagonist sequences exhibit cytostatic; immunosuppressive; neotropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; antitumor; antitumor; antitumor; antifungal; antiparasitic and cardiac activity. The polynucleotide and protein sequences are used in the diagnosis of cancer, particularly breast and ovarian cancer. The nucleic acid sequences, proteins, agonists and agonists may also be used in the diagnosis, prevention and treatment of immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC

CC cardiovascular disorders such as myocardial ischaemias; wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases
 XX
 SQ Sequence 1028 BP; 320 A; 174 C; 199 G; 335 T; 0 U; 0 Other;
 Query Match 1.8%; Score 99.6; DB 3; Length 1028;
 Best Local Similarity 70.8%; Pred. No. 1.5e-14;
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;
 QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAATGTGGCAACCCCTTATC--TAACCTGAA 5271
 DB 757 ACAGTAAAGTCCATTTCTGAGTGATGAATGTGAACACTCTTCATCTTTAACTTGA 816
 QY 5272 ATCAAAACCGTCARATTTTATTTTT-----TATAATTTAAGAAAGAGTGGGGAATGAC 5325
 DB 817 ATCAAACTATCAGATTTTATTTTTTGTATTAATTTAGGAAGTAAAGTTAGGGGACTAGA 876
 QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGTAG 5384
 DB 877 AGACTCTAAATTGGCTTCTACAGATCAATAATTTAAATGTAAC TAGTTGGGATTTATAG 936
 QY 5385 CTCAGATTATTTGTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGC 5444
 DB 937 TTAATAATTATTTGTGTATATAACATAACTAACTCTGTAAATTTGTAATAATATATTTC 996
 QY 5445 AATTATTAAA 5454
 DB 997 AATTATTAAA 1006
 RESULT 14
 AAH13104/c
 ID AAH13104 standard; cDNA; 482 BP.
 XX
 AC AAH13104;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:9939.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX
 PS Claim 3; SEQ ID NO 9939; 2537pp + Sequence Listing; English.
 XX
 XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 482 BP; 170 A; 83 C; 52 G; 170 T; 0 U; 7 Other;
 Query Match 1.8%; Score 98; DB 4; Length 482;
 Best Local Similarity 70.4%; Pred. No. 2.4e-14;
 Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;
 QY 5214 ACAGTGGGGCCACACTCTGAGGGCTGAATGTGGCAACCCCTTATC--TAACCTGAA 5271
 DB 261 ACAGTAAAGTCCATTTCTGAGTGATGAATGTGAACACTCTTCATCTTTAACTTGA 202
 QY 5272 ATCAAAACCGTCARATTTTATTTTT-----TATAATTTAAGAAAGAGTGGGGAATGAC 5325
 DB 201 ATCAAACTATCAGATTTTATTTTTTGTATTAATTTAGGAAGTAAAGTTAGGGGACTAGA 142
 QY 5326 ATTTTTCAGTTGGCTTTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGTAG 5384
 DB 141 AGACTCTAAATGGCTTCTACAGATCAATAATTTAAATGTAAC TAGTTGGGATTTATAG 82
 QY 5385 CTCAGATTATTTGTATATATAATTAATTAATTAATTAATTAATTAATTAATTAATTTGC 5444
 DB 81 TTAATAATTATTTGTGTATATAACATAATTAATTAATTAATTAATTAATTAATTTGC 22
 QY 5445 AATTATTAAA 5454
 DB 21 AATTATTAAA 12
 RESULT 15
 AAH16576
 ID AAH16576 standard; cDNA; 2148 BP.
 XX
 AC AAH16576;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:15655.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX

THIS PAGE BLANK (USPTO)

ALIGNMENTS

RESULT 1
US-09-513-999C-33878
; Sequence 33878, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 33878
; LENGTH: 218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-33878

Query Match 1.7%; Score 93.8; DB 4; Length 218;
Best Local Similarity 70.6%; Pred. No. 1.8e-15;
Matches 125; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 1550 ACAGAGCCCCCAGGAGCAGTCTGTTTGTGAATTACCGTGGAAGAAATGTGAAGA 1609
Db 42 AGAGAGCCCTTCTGAGGCGTGTCCATCTAGTGAGATACCTATGGAAGAGGCGCTGGAGA 101
QY 1610 GGACCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGAGCCCTTCTGTTTCTACCAAG 1669
Db 102 GGGCGAATAAGTGTAGTGTGATTACTATACCTAGAGGTGACCTTCCCATTTCTGCCAG 161
QY 1670 ACCAGTTGTAGCAACAACTGATAGATATATTTTGGAGGCGGCCCCAGTGAAGT 1726
Db 162 ACCAGCTTGTAGTACAACTGATAGATATATTTTGGAGGCGTGCATCCAGTGAAGT 218

RESULT 2
US-08-726-725-1/c
; Sequence 1, Application US/08726725
; Patent No. 5773290
; GENERAL INFORMATION:
; APPLICANT: Gould, Michael N.
; APPLICANT: Chen, Kai-Shun
; TITLE OF INVENTION: MAMMARY GLAND-SPECIFIC PROMOTERS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,725
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433

; REFERENCE/DOCKET NUMBER: 960296.93863
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3046 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-726-725-1

Query Match 1.4%; Score 78.8; DB 1; Length 3046;
Best Local Similarity 71.2%; Pred. No. 1.6e-10;
Matches 104; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 3613 TTTTCTCCCTCTTTTTTTTTTCCCCCAGAGCTGAGACCAACCCAGGCGTTTGCACCTT 3672
Db 421 TTTTCTTTTTTTTTTTTTTTTTTTCGGAGCTGGGACCGAACCCAGGCGCTTGGCGCTT 362
QY 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAACCAACCCCGCCCTTTCTTTT 3732
Db 361 CCTAGSTAAGCGCTCTGCCACTGAGCTAAATCCCAACCCCGCCCTTTCTTTTATTAT 302
QY 3733 TTTAAAGACATGGTCTTTATATAGTCT 3758
Db 301 TTATTATTATTGATATGAGGTCT 276

RESULT 3
US-10-329-668-5
; Sequence 5, Application US/10329668
; Patent No. 6696473
; GENERAL INFORMATION:
; APPLICANT: Martin Richard
; APPLICANT: Brenton Todd
; APPLICANT: Kahl Jeffrey Dean
; APPLICANT: Wang Tie-Lin
; TITLE OF INVENTION: HETEROCYCLIC MODULATORS OF NUCLEAR RECEPTORS
; FILE REFERENCE: 38205-3001
; CURRENT APPLICATION NUMBER: US/10/329,668
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/342,720
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2070
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (172)...(1581)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GeneBank U18374
; DATABASE ENTRY DATE: 1995-06-21
US-10-329-668-5

Query Match 1.4%; Score 75.2; DB 4; Length 2070;
Best Local Similarity 76.7%; Pred. No. 1.2e-09;
Matches 92; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 3593 TGTGTCACTTCAGTGGTGGATTTTCTCCCTCTTTTTTTTCCCCCAGAGCTGAGGACC 3652
Db 1896 TTATCAAAATAGCTGTTTCTCTCTTTTTTTTTTTTTTTTTCGGAGCTGGGACT 1955
QY 3653 AAACCCAGGCGTTTGCACCTTGTCTAGCAAGCGCTCTACCACTGAGCTAAATCCCAACCC 3712
Db 1956 GAACCCAGGCGCTTGGCTTGTCTAGCAAGCGCTCTACCACTGAGCTAAATCCCAACCC 2015

RESULT 4

RESULT 5
US-08-791-849A-14
Sequence 14, Application US/08791849A
Patent No. 5914449
GENERAL INFORMATION:
APPLICANT: Makoto MURASE et al.
TITLE OF INVENTION: Method for In-
Title of Invention: Lipid Content
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Pos-
Street: 805 Fifteenth Street, N
City: Washington
State: D.C.
Country: U.S.A.
Zip: 20005
COMPUTER READABLE FORM:
Medium Type: Diskette, 3.5 inch
Computer: IBM PC compatible
Operating System: MS-DOS
Software: Wordperfect 5.1
CURRENT APPLICATION DATA:
Application Number: US/08/791,8
Filing Date: January 30, 1997
Classification: 800
Prior Application Data:
Application Number:
Filing Date:
ATTORNEY/AGENT INFORMATION:
Name: Warren M. Cheek, Jr.
Registration Number: 33,367
Reference/Docket Number:
TELECOMMUNICATION INFORMATION:
Telephone: 202-371-8850
Telefax:
Telex:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:

RESULT 6
US-08-135-511-31/c
Sequence 31, Application US/08135511
Patent No. 5558999
GENERAL INFORMATION:
APPLICANT: Chiang, John
TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene
TITLE OF INVENTION: Regulatory Elements and Methods for Using Them
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135,511
FILING DATE: 13-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/175
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

Db 4888 GGGCTTGGCGTGTAGGCAAGTGTCTTACCCTAGCTAAATCCCGAGCCCGAGCACT 4829
Qy 3720 CCCCTTTT 3727
Db 4828 AAGCTTTT 4821
RESULT 9
US-09-600-319-3/c
; Sequence 3, Application US/09600319
; Patent No. 6780610
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary
; APPLICANT: Madsen, Cort
; TITLE OF INVENTION: Identification of a Smooth Muscle Cell (SMC) Specific Smooth Musc
; TITLE OF INVENTION: Myosin Heavy Chain (SM-MHC) Promoter/Enhancer
; FILE REFERENCE: 00241-03
; CURRENT APPLICATION NUMBER: US/09/600,319
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: PCT/US99/01038
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: 60/071,300
; PRIOR FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 16011
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-600-319-3
Query Match 1.3%; Score 71.2; DB 4; Length 16011;
Best Local Similarity 75.9%; Pred. No. 6.4e-08;
Matches 88; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
Qy 3612 ATTTTCTCCCTCTTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTGCACT 3671
Db 14537 ATTTTCTCCCTCTTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTGCGCT 14478
Qy 3672 TGCTAGGCAAGCGTCTTACCCTAGCTAAATCCCGAGCCCGAGCCCTTTT 3727
Db 14477 TCCTAGGTAGCGTCTTACCCTAGCTAAATCCCGAGCCCGAGCCCTTCATTT 14422
RESULT 10
US-09-600-319-3
; Sequence 3, Application US/09600319
; Patent No. 6780610
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary
; APPLICANT: Madsen, Cort
; TITLE OF INVENTION: Identification of a Smooth Muscle Cell (SMC) Specific Smooth Musc
; TITLE OF INVENTION: Myosin Heavy Chain (SM-MHC) Promoter/Enhancer
; FILE REFERENCE: 00241-03
; CURRENT APPLICATION NUMBER: US/09/600,319
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: PCT/US99/01038
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: 60/071,300
; PRIOR FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 16011
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-600-319-3
Query Match 1.3%; Score 69.8; DB 4; Length 16011;
Best Local Similarity 86.5%; Pred. No. 1.6e-07;
Matches 77; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 3624 TTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTTGCCTTGTAGGCAAGC 3683

Db 7641 TTTTTCCTCCCTCTTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTGCGCTTGTAGGCAAGC 7700
Qy 3684 GCTCTACCCTAGCTAAATCCCGCAAAACC 3712
Db 7701 GCTCTACCCTAGCTAAATCCCGCAAGCCC 7729
RESULT 11
US-08-974-022-1
; Sequence 1, Application US/08974022
; Patent No. 6015938
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,022
; FILING DATE: 12-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1326
; US-08-974-022-1
Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 3613 TTTTTCCTCCCTCTTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTGCACTT 3672
Db 1672 TTTTTCCTCCCTCTTTTTCCTCCCGAGCTGAGGACCAACCCAGGCTTTGCGCTT 1731
Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCGCAAAACCACCCCTTTTCTTTT 3732
Db 1732 GCGAGCAAGTGTCTTACCCTAGCTAAATCTCCAAACCCCTGAAGGCTCTTTCTTCT 1791
Qy 3733 TTTAAAGACATGTCTTATATAGCTAGGCTGCGCTTAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATTTCTTTTCTACAATTCGTATCA 1835
RESULT 12
US-08-795-445A-1
; Sequence 1, Application US/08795445A

```
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1326
; US-08-795-445A-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACTT 3672
Db 1672 TTTATTTATTTTATTTCTTTTTCGGAGCTGGGACCGAACCCAGGCGCTTGC 1731
Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAACCCAGGCGCTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGTCTACCACTGAGCTAAATCTCCAAACCCGCTTCTTTCTTCT 1791
Qy 3733 TTTAAAGACATGCTCTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATTTCTTTTCTACAAATTCGTATCA 1835

RESULT 13
US-08-795-447A-1
; Sequence 1, Application US/08795447A
; Patent No. 6284728
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: Osteoprotegerin
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: One Amgen Center Drive

; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,447A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378D2
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1326
; US-08-795-447A-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTCCTCCAGAGCTGAGGACCAACCCAGGCGCTTTGCACTT 3672
Db 1672 TTTATTTATTTTATTTCTTTTTCGGAGCTGGGACCGAACCCAGGCGCTTGC 1731
Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAACCCAGGCGCTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGTCTACCACTGAGCTAAATCTCCAAACCCGCTTCTTTCTTCT 1791
Qy 3733 TTTAAAGACATGCTCTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATTTCTTTTCTACAAATTCGTATCA 1835

RESULT 14
US-08-974-186-1
; Sequence 1, Application US/08974186
; Patent No. 6284740
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,186
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```


APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-974-186-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTTTTTTCCCGCAGAGCTGAGGACCAACCCAGGGCTTTGCACTT 3672
Db 1672 TTTATTATTTTATTCTTTTTTTCGGAGCTGGGACCGAACCCAGGGCTTGGCGTT 1731

Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAAAACCAACCCCGCCCTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGCTCTACCACTGAGCTAAATCTCCAACCCCTGAAGGCTCTTCTTTCT 1791

Qy 3733 TTTAAAGACATGGTCTTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATTTCTTTTCTACAATTTCGTATCA 1835

RESULT 15
US-08-795-446B-1
Sequence 1, Application US/08795446B
Patent No. 628032
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-795-446B-1

Query Match 1.3%; Score 69.6; DB 3; Length 2432;
Best Local Similarity 64.0%; Pred. No. 4.8e-08;
Matches 105; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 3613 TTTTCTCCCTCTTTTTTTTTTCCCGCAGAGCTGAGGACCAACCCAGGGCTTTGCACTT 3672
Db 1672 TTTATTATTTTATTCTTTTTTTCGGAGCTGGGACCGAACCCAGGGCTTGGCGTT 1731

Qy 3673 GCTAGGCAAGCGCTCTACCACTGAGCTAAATCCCAAAACCAACCCCGCCCTTTCTTTT 3732
Db 1732 GCGAGGCAAGTGCTCTACCACTGAGCTAAATCTCCAACCCCTGAAGGCTCTTCTTTCT 1791

Qy 3733 TTTAAAGACATGGTCTTATATAGTCTAGGCTGGCGCTTAAACTCA 3776
Db 1792 GCCTCTGATAGTCTATGACATTTCTTTTCTACAATTTCGTATCA 1835

Search completed: September 16, 2005, 09:17:08
Job time : 827.974 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 15:05:39 ; Search time 3234.64 Seconds
(without alignments)
11262.759 Million cell updates/sec

Title: US-10-650-482-3

Perfect score: 5468

Sequence: 1 cggctccgtctgcctg.....attaaaaaaaaaaaaa 5468

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7389322 seqs, 3331285599 residues

Total number of hits satisfying chosen parameters: 14778644

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
16:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
18:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq*
19:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq*
20:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq*
21:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq*
22:	/cgn2_6/ptodata/1/pubpna/US10J_NEW_PUB.seq*
23:	/cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq*
24:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
25:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
26:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5468	100.0	5468	19	US-10-650-482-3
2	1324	24.2	2942	19	US-10-650-482-1
3	1189.2	21.7	2324	9	US-09-925-299-168
4	1189.2	21.7	2324	10	US-09-925-299-168
5	542.6	9.9	883	18	US-10-236-115-224
6	215.8	3.9	465	22	US-10-756-149-3997
7	181.6	3.3	540	9	US-09-917-800A-1296

8	170.8	3.1	399	9	US-09-983-965-2096	Sequence 2096, Ap
9	122	2.2	242	9	US-09-783-590-12337	Sequence 12337, A
10	109	2.0	421	10	US-09-918-995-36363	Sequence 36363, A
11	99.6	1.8	1028	9	US-09-925-298-148	Sequence 148, App
12	99.6	1.8	1028	14	US-10-102-806-148	Sequence 148, App
13	98	1.8	2148	17	US-10-170-385-162	Sequence 162, App
14	98	1.8	2148	17	US-10-131-827-8849	Sequence 8849, Ap
15	79.4	1.5	8351	9	US-09-917-800A-1399	Sequence 1399, Ap
16	79.4	1.5	8351	17	US-10-388-934-665	Sequence 665, App
17	79.4	1.5	8351	17	US-10-388-934-719	Sequence 719, App
18	79.4	1.5	8351	17	US-10-191-803-1094	Sequence 1094, Ap
19	79.4	1.5	8351	18	US-10-152-319A-1461	Sequence 1461, Ap
20	79.4	1.5	119414	13	US-10-087-192-1807	Sequence 1807, Ap
c 21	79.2	1.4	1448	18	US-10-152-319A-2090	Sequence 2090, Ap
c 22	79.2	1.4	4529	17	US-10-191-803-273	Sequence 273, App
23	79.2	1.4	4529	18	US-10-152-319A-2052	Sequence 2052, Ap
24	79.2	1.4	25891	13	US-10-087-192-1657	Sequence 1657, Ap
c 25	79.2	1.4	53963	18	US-10-052-482-43	Sequence 43, Appl
c 26	78.6	1.4	1496	10	US-09-866-050A-519	Sequence 519, App
c 27	78.6	1.4	1496	14	US-10-152-661-519	Sequence 519, App
c 28	78.2	1.4	71132	13	US-10-087-192-1867	Sequence 1867, Ap
c 29	77.6	1.4	5531	17	US-10-191-803-234	Sequence 234, App
c 30	77.6	1.4	5531	18	US-10-152-319A-1974	Sequence 1974, Ap
c 31	77.6	1.4	215980	10	US-09-972-546-16	Sequence 16, Appl
c 32	77.6	1.4	215980	21	US-10-735-256-16	Sequence 16, Appl
c 33	77.2	1.4	36119	19	US-10-087-192-97	Sequence 97, Appl
c 34	77	1.4	6251	9	US-09-917-800A-472	Sequence 472, App
35	76.6	1.4	6251	18	US-10-152-319A-527	Sequence 527, App
36	76.6	1.4	479	17	US-10-388-934-727	Sequence 727, App
c 37	76	1.4	552	21	US-10-764-420-317	Sequence 317, App
c 38	76	1.4	575	21	US-10-486-708-410	Sequence 410, App
40	76	1.4	120239	19	US-10-322-281-271	Sequence 271, App
41	75.4	1.4	508	9	US-09-917-800A-319	Sequence 319, App
c 42	75.4	1.4	21688	21	US-10-621-911A-9	Sequence 9, Appl1
c 43	75.4	1.4	21688	21	US-10-646-390A-9	Sequence 9, Appl1
c 44	75.2	1.4	2070	16	US-10-327-813-5	Sequence 5, Appl1
45	75.2	1.4	2070	17	US-10-329-668-5	Sequence 5, Appl1
46	75.2	1.4	2070	18	US-10-152-319A-1756	Sequence 1756, Ap
47	75.2	1.4	2070	19	US-10-717-049-5	Sequence 5, Appl1
48	75.2	1.4	2070	21	US-10-899-458-13	Sequence 13, Appl1
c 49	75	1.4	2472	17	US-10-408-713-5	Sequence 5, Appl1
c 50	75	1.4	2485	18	US-10-152-319A-1901	Sequence 1901, Ap
c 51	74.8	1.4	2010	9	US-09-809-545A-52	Sequence 52, Appl1
52	74.6	1.4	227	10	US-09-930-213-739	Sequence 739, App
c 53	74.4	1.4	13553	17	US-10-388-934-208	Sequence 208, App
c 54	74.2	1.4	13011	17	US-10-388-934-36	Sequence 36, Appl1
55	74.2	1.4	13011	18	US-10-152-319A-2150	Sequence 2150, Ap
c 56	74	1.4	3134	10	US-09-866-050A-478	Sequence 478, App
c 57	74	1.4	3134	14	US-10-152-661-478	Sequence 27, Appl1
58	73	1.3	109565	19	US-10-322-281-27	Sequence 27, Appl1
59	72.8	1.3	1089	20	US-10-466-634-3	Sequence 3, Appl1
60	72.8	1.3	1686	17	US-10-388-934-549	Sequence 549, App
c 61	72.8	1.3	2066	18	US-10-152-319A-1553	Sequence 1553, Ap
c 62	72.8	1.3	3561	21	US-10-486-708-210	Sequence 210, App
c 63	72.8	1.3	4964	17	US-10-388-934-565	Sequence 565, App
c 64	72.6	1.3	175338	13	US-10-087-192-1861	Sequence 1861, Ap
c 65	72.4	1.3	864	18	US-10-240-240A-63	Sequence 63, Appl1
c 66	72.4	1.3	4039	17	US-10-336-091-31	Sequence 31, Appl1
c 67	72.4	1.3	189158	13	US-10-087-192-415	Sequence 415, App
c 68	72.4	1.3	192673	19	US-10-331-053-1	Sequence 1, Appl1
c 69	72.2	1.3	23713	21	US-10-936-273-13	Sequence 13, Appl1
c 70	72.2	1.3	32963	13	US-10-087-192-1801	Sequence 1801, Ap
c 71	72.2	1.3	107304	19	US-10-322-281-249	Sequence 249, App
72	72	1.3	165961	17	US-10-085-117-241	Sequence 241, App
c 73	71.8	1.3	983	18	US-10-240-240A-85	Sequence 85, Appl1
74	71.8	1.3	1314	18	US-10-152-319A-1487	Sequence 1487, Ap
75	71.8	1.3	4447	21	US-10-486-706-374	Sequence 374, App
76	71.8	1.3	7656	21	US-10-621-911A-7	Sequence 7, Appl1
77	71.8	1.3	7656	21	US-10-646-390A-7	Sequence 7, Appl1
78	71.8	1.3	12460	14	US-10-175-523-93	Sequence 93, Appl1
79	71.8	1.3	12460	24	US-11-099-266-93	Sequence 93, Appl1
80	71.8	1.3	15316	21	US-10-889-101-118	Sequence 118, App

1441 GGATGGACCCCAACATTGCAAGATTAACCCAGCAGAGCGGGTGTCCCTGCTGCAGACA 1500
1441 GGATGGACCCCAACATTGCAAGATTAACCCAGCAGAGCGGGTGTCCCTGCTGCAGACA 1500
1501 GGGCGGAGCCCACTGAGAGAAAACAGAAATGGTGTATCAAGAAATTTTCAAGAGCCGCC 1560
1501 GGGCGGAGCCCACTGAGAGAAAACAGAAATGGTGTATCAAGAAATTTTCAAGAGCCGCC 1560
1561 AGGGAAGCAGTCTGTTTGTGAAATTTACCCGTGGGAAAAGAAATGTAAGAGAGACCACTA 1620
1561 AGGGAAGCAGTCTGTTTGTGAAATTTACCCGTGGGAAAAGAAATGTAAGAGAGACCACTA 1620
1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCCCTGTTTCTACAGAGCCAGTGTGTA 1680
1621 ATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCCCTGTTTCTACAGAGCCAGTGTGTA 1680
1681 GCAACAACTGATAGATTAATTTTGGAGAGCGGCCCGCCAGTGACTTGGAAAGCCAGCTCTG 1740
1681 GCAACAACTGATAGATTAATTTTGGAGAGCGGCCCGCCAGTGACTTGGAAAGCCAGCTCTG 1740
1741 ATCTGAAAGTGAGGATTTGGGCGGAGGAACCTGAGGACGATGGCTTTTGATAGCGATGGCT 1800
1741 ATCTGAAAGTGAGGATTTGGGCGGAGGAACCTGAGGACGATGGCTTTTGATAGCGATGGCT 1800
1801 CCCTGTCTGATCAGACGTGGAACAGGACTCGGAAGGCCCTTCACTTTTGGAACTCTTTCC 1860
1801 CCCTGTCTGATCAGACGTGGAACAGGACTCGGAAGGCCCTTCACTTTTGGAACTCTTTCC 1860
1861 ACAGTGTAGATCTTTAACAACCCCAAACTTTACAGCCACGATTCAGACGCGCTGCCAGAA 1920
1861 ACAGTGTAGATCTTTAACAACCCCAAACTTTACAGCCACGATTCAGACGCGCTGCCAGAA 1920
1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGTCTGGCAGCTGTGGTGTAGGGA 1980
1921 TTGCCCCCAGAGACCCATCAGATTCAGGACATCTGTCTGGCAGCTGTGGTGTAGGGA 1980
1981 GCTGTCAAGGAGGACCCCTTCCGAGACCCCGACCATAGTTCCGGGGAGGAAGTGA 2040
1981 GCTGTCAAGGAGGACCCCTTCCGAGACCCCGACCATAGTTCCGGGGAGGAAGTGA 2040
2041 GGGAAACGAGTGCAGATGAAGCAGAGATCTTAAATTTGTGGAATCTTTCTGTCAATCTG 2100
2041 GGGAAACGAGTGCAGATGAAGCAGAGATCTTAAATTTGTGGAATCTTTCTGTCAATCTG 2100
2101 AGGACCCCTCAACCTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAAGGGAAGAAATGGA 2160
2101 AGGACCCCTCAACCTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAAGGGAAGAAATGGA 2160
2161 AAGGCCGTCAAGACTCAAAAGCCCTCTTCTGAGGTCAAGTGGCCCTTCTCGGCCATCATA 2220
2161 AAGGCCGTCAAGACTCAAAAGCCCTCTTCTGAGGTCAAGTGGCCCTTCTCGGCCATCATA 2220
2221 CCTTACTTTCTGTAAGGCCAGCTGTTTGAAGAGGAGATACCCATATCAAGAGAAAAGGTAACCT 2280
2221 CCTTACTTTCTGTAAGGCCAGCTGTTTGAAGAGGAGATACCCATATCAAGAGAAAAGGTAACCT 2280
2281 GGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACCCATATCAAGAGAAAAGGTAACCT 2340
2281 GGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACCCATATCAAGAGAAAAGGTAACCT 2340
2341 TCCTGGGAAGAGTACTGAGTATTAATAAGTGGTGTAGGATCGCAAGAGCCATGGG 2400
2341 TCCTGGGAAGAGTACTGAGTATTAATAAGTGGTGTAGGATCGCAAGAGCCATGGG 2400
2401 AAGAAATTTGCAAGGATGGATGCAAGTTCCAGAAACGAAATTCAGAAACAGAAATTTGCCA 2460
2401 AAGAAATTTGCAAGGATGGATGCAAGTTCCAGAAACGAAATTCAGAAACAGAAATTTGCCA 2460
2461 TTGGCTACTGCTTGGCCCTTTGAGCAGCAGAGAAATGTTTAAATAGACTGAGGATCGAGT 2520
2461 TTGGCTACTGCTTGGCCCTTTGAGCAGCAGAGAAATGTTTAAATAGACTGAGGATCGAGT 2520

2521 CAAAAGGACTTACTGTGTGACAGCAATGTTTAAAGAGTGAACAGCCTGCAACCCGCTGCCAC 2580
2521 CAAAAGGACTTACTGTGTGACAGCAATGTTTAAAGAGTGAACAGCCTGCAACCCGCTGCCAC 2580
2581 TCTGTCTCTTACTTGTAGAGTTCCTTTAAAAAACAACACTGSCAGCTGTCTTGGACATG 2640
2581 TCTGTCTCTTACTTGTAGAGTTCCTTTAAAAAACAACACTGSCAGCTGTCTTGGACATG 2640
2641 TTTTAAAGAAAAACAATCTGTATCTAGAGATGCAAGTTGATTATTTTGGGTAAATGTGTCT 2700
2641 TTTTAAAGAAAAACAATCTGTATCTAGAGATGCAAGTTGATTATTTTGGGTAAATGTGTCT 2700
2701 CATTAAGAAACCAACCTCCGATTAATGAAGAAATCTCTTATCTGTAATCTCTTTTCCCTA 2760
2701 CATTAAGAAACCAACCTCCGATTAATGAAGAAATCTCTTATCTGTAATCTCTTTTCCCTA 2760
2761 TTTTAAAGTGGATGAGTGGTGTCTTTTGAAGAGGCTCTCACTCAATAATTTCTGTTTGGCC 2820
2761 TTTTAAAGTGGATGAGTGGTGTCTTTTGAAGAGGCTCTCACTCAATAATTTCTGTTTGGCC 2820
2821 TAGGTCTATGATGATAGACCCAGGTTCGCTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880
2821 TAGGTCTATGATGATAGACCCAGGTTCGCTTGAAGTTGGTATTTTCTTGGCTCTGTTTCC 2880
2881 TGTGCCACCAATGCCCAGCTGGAAGTGTCTTTTAAATATCTCTGCTTACCTGGGGTGAGAG 2940
2881 TGTGCCACCAATGCCCAGCTGGAAGTGTCTTTTAAATATCTCTGCTTACCTGGGGTGAGAG 2940
2941 TTTTAAATTTGCACTTTTCAAAACGTTTCTTCTGAGGCGAGGGCTGTGTTTACATAC 3000
2941 TTTTAAATTTGCACTTTTCAAAACGTTTCTTCTGAGGCGAGGGCTGTGTTTACATAC 3000
3001 AGGCTTCAGGCCAGCTAGGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAACAGAAACA 3060
3001 AGGCTTCAGGCCAGCTAGGGCTATGTGGTGAGACCTAGTCTTAGAGGACAAACAGAAACA 3060
3061 AACAGTCAGGTATCTGTGGAACCTGAGGCGAGGAGTAGGAAGTCCAGGTATGCTTGGGA 3120
3061 AACAGTCAGGTATCTGTGGAACCTGAGGCGAGGAGTAGGAAGTCCAGGTATGCTTGGGA 3120
3121 ACTCAGTGAACCTGCAACTCAAGCCAGCATGGGAGTTTAGCAAGACCTTTTACACTC 3180
3121 ACTCAGTGAACCTGCAACTCAAGCCAGCATGGGAGTTTAGCAAGACCTTTTACACTC 3180
3181 AGAATGGAAGGAAGCTGGGATATAGCTCGGTAGCAGAGGCCCTTCCCTACATGTGCA 3240
3181 AGAATGGAAGGAAGCTGGGATATAGCTCGGTAGCAGAGGCCCTTCCCTACATGTGCA 3240
3241 AGGTCTCTGGTTTCAAGTCCCGAGTACCTGCAAAATAGAAAGAAAAAACAATTTGGAATAAC 3300
3241 AGGTCTCTGGTTTCAAGTCCCGAGTACCTGCAAAATAGAAAGAAAAAACAATTTGGAATAAC 3300
3301 TATAAGGTTTAAAGCCATAGTCAAGTCTTAACTCAAAATATGATGATGATGATGATGATGATG 3360
3301 TATAAGGTTTAAAGCCATAGTCAAGTCTTAACTCAAAATATGATGATGATGATGATGATGATG 3360
3361 GTGCTTTTCTGTTCTTAAATTAATGGGCTTTATGGGTTTATGTTTGTATGTTTGTATG 3420
3361 GTGCTTTTCTGTTCTTAAATTAATGGGCTTTATGGGTTTATGTTTGTATGTTTGTATG 3420
3421 TCTGTATATTTTGAAGTGGGATGAGCAGGGTCTCACTGTGGCTTGGAGCTATGG 3480
3421 TCTGTATATTTTGAAGTGGGATGAGCAGGGTCTCACTGTGGCTTGGAGCTATGG 3480
3481 CAGTTTCTGCTTCAACTGAGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTCA 3540
3481 CAGTTTCTGCTTCAACTGAGTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTCA 3540
3541 TAGAATTTGACTTAATGAAGGCAAAACCTGTTATCAACCTTAAAGACATTTGATGTGTAC 3600
3541 TAGAATTTGACTTAATGAAGGCAAAACCTGTTATCAACCTTAAAGACATTTGATGTGTAC 3600
3601 TTCAGTGGTGAATTTTCTCCCTCTCTTTTTCCTCCAGAGCTGAGGACCAAAACCCAG 3660

; TYPE: DNA		Query Match		24.28; Score 1324; DB 19; Length 2942;	
; ORGANISM: Homo sapiens		Best Local Similarity		72.58; Pred. No. 0;	
US-10-650-482-1		Matches 1964; Conservative		0; Mismatches 660; Indels 84; Gaps 16;	
Qy	151	CCGTGCGCGCGCGGAGGAGGCTCTTCTCTATGTTGGAGCGATCTCACACGCGCTAGG	210	TTTCAGTGGTATCCTATTTGCTGAACCCCTTCTACCTGGACTGCTTCTCCTAGGTAGA	1101
Db	96	CCATGCGCGAGCGCGTGGGGGACTCTCTATCCATGGTGTGAAGCGTCGAGCGGACTAGG	155	TTTCAGTGGTATCCTATTTGCTGAACCCCTTCTACCTGGACTGCTTCTCCTAGGTAGA	1110
Qy	211	ACGTCTCTCTCTCTAGCCGGATGGACCTAACCGCGGTGCGCACCGCTTTCGGCGGGCTC	270	GCTGCGCTGTACAGACAGCGCTGGAGGTGGCCAGTGTGGGGTTCCTCGAACACATAACCC	1210
Db	156	GAACCTCTCTCTCTCTAGCCGGATGGACCTAACCGCGGTGCGCACCGCTTTCGGCGGGCTC	214	AGTCAGCTATCAGAACAGTATAGCAGGTAGTGGCTTCCAGACACTTAACCC	1170
Qy	271	TGGGCGGTGCGGTGCGACACTCTGTTGGGAAGCGCGCTCTCTGGGCTCTCTCTGCGG	330	AGAGAGCTGTATCTTTCTGAAGATGGTGTACCTCAGCCGTTGCGGGCAGAGATGTC	1270
Db	215	TCCTCCCTGTGTTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTTGGCTTCTGCTGG	274	AGAGAGCAGTGTCTGAGAGAGACCATTTGTATCCAGCGCTGAGTGCAGAACTCAT	1230
Qy	331	CGCGGGAATCGGACTGCAGTACCCACTCCGTGGCTGGGCAAGCGGAGACTGTGTAGAC	390	GGCAACCGCTGGAGAAGGTGTCCGCTCTCTACAGAGGCTGCCGGAATCCACCA	1330
Db	275	CTCGAAGATCGGCTGAGCAGCGACCGCACCGCTGGGCAAGCGGAGACTGTGTAGCT	334	TCGCGCTCTGTGGCAGGGATGTCACCTCTTCTACGGAAGGCTACCCAGAAATCACC	1290
Qy	391	TCGGATCCAGCTGCGCTGACCGCGTGTAGCTGTCTCTCTCTCTGTCTGAGAGCGGCC	450	CTTCTGTATGAACCGGCTAGAAATTCCT---CCAGGCTAACAAAGGCAAGATTACCC	1387
Db	335	TCCTCCGAATCCGCTGACCTCCAGCGCTGAGCGCGCGGCTCTACCTGAGAGACTGTC	394	TCCTCCGATGATTAATGGCTATCATAGCTGGAGGAGGAACAATACTTCTCCGGATGGA	1447
Qy	451	AAG-GAAGAGATGGAGACAGGACGACGACGAGCGCGGAGCGGCTGCGCTGCGCTG	509	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	395	AGAAAAAAGAGATGGAGCGCGGAGACAGCGGATTCGCGGAACCGCTTGGCCCTCGGG	454	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	510	GGCTCTCTGTTTGGGCTGCGC---TTCTCTTGGCGATTCGACCGCTCTCTCTGGAGTTC	566	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	455	GGCTCTCTGTTTGGGCTGCGC---TTCTCTTGGCGATTCGACCGCTCTCTCTGGAGTTC	514	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	567	CGCGCGCTCTCTCTCGCAAAATCCCGGAATCTC-----GCTCTGCCC	611	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	515	CCGACGCTCTTGGCGCGGAAACTCCGGAACCCACACTGCTCTCTGCGCCAGCCC	574	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	612	GAGCGTGGACAGTACTCGACCAATTCGTTTCTCAGCTCTCTGCGCTGCTCTCTCTG	671	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	575	GAGACTCGGTCAGTACTGACGAAATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	634	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	672	CTATTCCAGAGCTGCTGCTTTTGAGCCAGCTTTTCCGGGGCGCTGATTTCTTACAGATGG	731	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	635	TTGCTTCAGAAAGTGCTAAATTTGGAGCCAACTTTCCGTGGATGTTTCCGACCAAGTGG	694	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	732	CTAGATTTTCCGCAAGTTACAGCGCTGAGAGCTTCGAGAGGACGGAGGAATCTGAC	791	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	695	CTAGATTTTCCGCAAGTTACAGCGCTGAGAGCTTCGAGAGGACGGAGGAATCTGAC	754	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	792	GCTCCACGCTGAGAGTCTCTGAGTTACCTGCGGCTGGACTCTTCCGGAAGACTCGC	851	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	755	GCCCCACAGCGCAAGAAATCTTTGAGTT-CGCTGCACTGCACTCTCTCTCTCTCTCT	810	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	852	GTCGTGAGTACTCTGCAATGGCTTAGAGGGGACTCCAGTGGCAGTCTGCTCTCTCAGA	911	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	811	GGTCAACAGTCTCTGATTTGGCTTAGAGGGGATCCACTGGCAATACTCTGCCCCCAGA	870	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	912	-CTGGAGTTAACTCAAGGCCCAAGGAAGCTTTTAGACTCTGACGCGGCCACTTTTCT	970	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	871	CCTAAATTTGAGCTTAAGGCCAAGGAAGTCTTTTGGACCCCTGACACAGGCTTTCT	930	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	971	CCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCCCTTCAAGTGTCTAGT	1030	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	931	CTTAGAGCAGCAGCTGTGGGAGTGGAGTGTGTGCCAGTAGCCCTTCAATCCCGCTCTGA	990	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	1031	CTCCACCGAGAACTTGACTCTTCACTCTGCGCTCTTGAGCGTTCAGAGCTTAGGTAA	1090	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350
Db	991	CTCTAACCGGAACTTGCTCTTTCGCTCTTGGGCTCTTAAACATTCAGCGATAGACAA	1050	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1387
Qy	1091	TTTCAAGGTAGTTTCTCTATCTCTGAAACCCCTTCTTACCTGGACTACCTTCCCCAGTTAGG	1150	CCCTGACAGGATTAATGGCTTACCAAGCTTCCCTCAACAGGCTACAAAGGCAAGATTAC	1350

Db 2131 TCAAAATCAGGGGAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCAT 2190
 Qy 2198 AGTGGCCCTCTCTGCGCATCATACCTTACTTCTTGTAAAGGCCAGCTGTTAGAGAGCCA 2257
 Db 2191 TGTGGCCATTTCTGAGTGTCAACCTTACTTCTTGTAAAGTGCGAGCTGTTGGGAGCCA 2250
 Qy 2258 AGAAGATAAATGTCCAGGCTGTGGGCTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCCA 2317
 Db 2251 AGAAGATGAATGTCCAGACTCGGTACAGGTGACGCTTCTTCTGGAGAGACACACACA 2310
 Qy 2318 TATCAAGAGAAAAAGTAACTCTCTCGAAGAGTACTGAGTATTATATAAGTGTGA 2377
 Db 2311 TGTCAAAAGAAAAAGTAACTCTCTCGAAGAGTACTGAGTATTATATAAGTGTGA 2370
 Qy 2378 TGAGATCGCAAGACCATGGGAGAAATTTGCAAGGGATGGATCGAGTTCGAGAACG 2437
 Db 2371 TGAGATCGCAAGACCATGGGAGAAATTTGCAAGGGATGGATCGAGTTCGAGAACG 2430
 Qy 2438 AATTCAGAAAAAGAAAGTTGGCTACTGTCTGGCTTTGAGCACAGAGAAAAAAT 2497
 Db 2431 AATTCAGAAAAAGAAAGTGTATTTGGATATTTGATATTTGACATTTGAAACACAGAGAAAGAT 2490
 Qy 2498 GTTTAATAGACT----GAGGATCGAGTCAAAAGGACTTACTGTGTACAGCAATGTTAAGA 2553
 Db 2491 GTTTAATAGACTCCAGGGAAACATGCTTCAAAAGGACTTAATGTCTCAAGCAATGTTGAGT 2550
 Qy 2554 AGTGA-ACAGCTGCAACCGGTGCCACTCTGTCTTACTTACTGAGA-GTTTCCCTTAAAA 2611
 Db 2551 TGGCAGCTGTAGTCTTAGCTAGCATACACTACCTCTTACCTGAGAGGTGCTTTTAAAA 2610
 Qy 2612 ACAAACTACGCGAGTGTCTTGGACATGTTTAAAGAAACAACTTGTATCTAGAGATG 2671
 Db 2611 ACNAATCTGGCAGCTGCTTGGACATTTTGGACATTTTGTAGGAAATGTAATTTGGATCT 2670
 Qy 2672 CAGTTTGAATTTTGGGTAAAGTGTCTCAATAGAA-----ACACCAACT 2717
 Db 2671 AGTTTAAATTTTGTGCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC 2730
 Qy 2718 CCGATAATGAAGAT-----CTCTATCTGTAACTCTCTTCTTCTTATTTAGTTGGAT 2771
 Db 2731 CTGATAATGAAGGATGAATAGTGTGATTTCTAATCTCTCCCTTTTGTGATTTAGTTGGAT 2790
 Qy 2772 GTGGGTTT 2779
 Db 2791 GTGCTTTT 2798

RESULT 3

US-09-925-299-168
 ; Sequence 168, Application US/09925299
 ; Patent No. US20020055627A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: P102
 ; CURRENT APPLICATION NUMBER: US/09/925,299
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 168
 ; LENGTH: 2324
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-299-168

Query Match 21.7%; Score 1189.2; DB 9; Length 2324;

Best Local Similarity 75.3%; Pred. No. 0;

Matches 1642; Conservative 4; Mismatches 477; Indels 59; Gaps 11;

Qy 519 TTCCGGCTGCCCTTCTTCGCGCATCGACGCTGCTCTTTCGGAGTTCGCCGCCCTTCC 578
 Db 43 TTTTCAACCTCTGTGCTGCTGCGCATCGAAGAGGCTCTTCTAAGTTTCCCGACGCTCTT 102
 Qy 579 TCTCGACAAATATCCCGGAACTTCGCTCT-----GCCGAGCGTTCGAGCC 623
 Db 103 GGCCCGGAAAACTCCGGGAAACCCACACTGCTTCTCTGCCACGCGCGAGACTCGGGTC 162
 Qy 624 AGGTACTGGAACCAATTTCTCAGCTCTCTTGCCTGCTCCCTAGCTATTTCCAGAG 683
 Db 163 AGTTACTGGAACCAATTTCTCAGCTCTCTTGCCTGCTCCCGGATTTGCTTTCAGAG 222
 Qy 684 CTGCTGCTTTTGGAGCCAGCTTTTCCGGGGCTGATTTCTTACAGATGGCTAGATTTTGGC 743
 Db 223 GTGCTAATTTTGGAGCCAACTTTTCCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTGGT 282
 Qy 744 GCAAGTTTACAGCGCTTTCGAGAGCTTCGAGAGAGCGGAGGAAATCTGACGCTCCACCGTG 803
 Db 283 GGAGTCTACAGCGCTTTCGAGAGCGCTTCGAGAGAGCGGAGGAAACCGAGCGCGCCACAGCG 342
 Qy 804 CAGAAGTCTCTGAGTTTACACTCGGCGCTGGAATCTTTCGCGAAGACTCGCGTCTCAGTACT 863
 Db 343 CAGAAATCTTTTGAATTT--CGCTGCAAGCTTCGACTCTCTC---AGACCTCTCGGTCAACAGTCC 398
 Qy 864 CTTGCAATTTGGCTAGAGAGGAGCTCCAGTGGCGATGCTCGTCTCTCAGA-CTGGAAGTTAA 922
 Db 399 CTTGATTTGGCTAGAGAGGAGGATCCACTGSCAATACTCGCCGCCACAGACCTTAAATTTGGA 458
 Qy 923 ACTCAAGGCCCAGAAAGAGCTTTAGACTCTGCAAGCGCCCACTTTCTCTCTGAGAGCAGA 982
 Db 459 GCTTAAGGCCCAAGGAAAGTGTCTTTGGACCTTCGAGCACAAGGCTTTTCTCTTAGAGCAGA 518
 Qy 983 GCTGTGGGGAGTGGAGTTGCTGCCAGTACGCTTCAAGCTGTCTAGTCTCTCCCAACCCAGA 1042
 Db 519 GCTGTGGGGAGTGGAGTTGCTGCCAGTACGCTTCAATCCGCTCTGACTCTAATACCCGGA 578
 Qy 1043 ACTTGAATCTTCAATCTCTGCGGCTCTGAGCGTTTCAAGCTTTAGGTAAATTTCAAGTAGT 1102
 Db 579 ACTTGGCTCTTCGCGCTCTGCGCTTTCTAAACATTTCAACGATAGACAAATTTCAAGTGTGT 638
 Qy 1103 TTCTATCTCTGAAACCTTCTTACTGAGACTACTTCCCGAGTTCAGGCTGAGGCTGCGCTGCA 1162
 Db 639 ATCTATTTGCTGAACCTTCTTCACTCTGAGCTGCTTTCTTAGGCTAGAGTCAAGCTATCA 698
 Qy 1163 GAGCAGCGCTGGAGTGGCCAGTTTGTGGGTTTCCGAACTAATACCCACAGAGACTGCTA 1222
 Db 699 GAACAGTGTAGAAATAGCGAGTGTGCGCTTCCAGACTAATACCCACAGAGAGAGCTG 758
 Qy 1223 TCTTTCTGAAGATGGTTGTCACTCTGAGCGTTTGGGGCAGAGATGTCGGCAACCGCTG 1282
 Db 759 CTTGAGAGAGGAGCCATTTGTCTATCCCGAGCGGTGARTGCAGAACTCATTTCCGSGCTCGTG 818
 Qy 1283 GAGNAGTGTCCGCTCTCTTACAGAGGCGCTTCCCGGAAATCCACCACTTCTGATGAA 1342
 Db 819 GCAGGGATGTCCACCTCTTCTATCGGAAAGGCTTACCGAAGATTTCCACCATCTTCGCATGAA 878
 Qy 1343 ACGGCTAGAAATTTCT---CCAGGCTTAAACAAAGGCAAGATTTACCCACCTCTGACCAAGA 1399
 Db 879 ACGCTGGAATTTCTTCAACAGGCTAGCAAGGGGCAAGATTTACCCACCTCTGACCAAGA 938
 Qy 1400 TAATGGCTATCATAGCTTGGAGGAGAAATTAACCTTTCTCCGGATGGACCCACCAACATTTG 1459
 Db 939 TAATGGCTATCCACAGCTGCGAGGAGGAAACACAGCTTCTCCGGATGGATCCAAACACTG 998
 Qy 1460 CACAGATAACCCACAGAGCGGCTGCTCTGCTGCGAGACAGGCC-----GGA 1507
 Db 999 CAGAGATAACCCACACAGATTTGTCTGCTGCTGGAGACATTTCTCTGGAACACCCAGGA 1058
 Qy 1508 GCCCACTGAGAAAAAACAGAAATTTGGTGAATTTCAAGAAAGTTTCAAC-----A 1552
 Db 1059 ATCCACTGAGAAAAAATAGAAATTTAATTAATCAAGAGTTTCCACTGCTTTTGGAGAGAA 1118
 Qy 1553 GAGCCCCCAGGGAGAGCAGTCTGTTTGTGAATTTACCGGTGGAAAAAAGAAATGTGAGAGGA 1612


```
Db 1119 GAGCCCTTCTGAGGCGTGTCCATCTAGTGTGATACCTATGGAAGGAGCGCTGGAGAGG 1178
Qy 1613 CCACACTAATCAACTGACCTCTCAGATAGAGGAGAGCGCTTCTGTTTCTACAGACC 1672
Db 1179 CCGAATAAGTGTAGTGTACTATCATCTAGAAAGGTGACCTTCCATTTCTGCAGACC 1238
Qy 1673 AGTTTGTAGCAACAACTGATAGATATATTTTGGGAGGCGCCCGACGTGACCTTGGAGC 1732
Db 1239 AGCTTGTAGTAACAACTGATAGATATATTTTGGGAGGTGCATCCAGTGACCTGGAAAC 1298
Qy 1733 CAGCTCTGATTTCTGAAGTGAAGATTGGGGCGAGAACCTGTGAGACACATGGCTTTGTATG 1792
Db 1299 AAGTTCTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGTGAGGATGATGTTTGTATG 1358
Qy 1793 CGATGGCTCCCTGTCTGAATCAGACGTTGGAACAGGACTCGGAAGGCTTCACTTTTGAA 1852
Db 1359 TGATAGCTACTGTGAGACTCAGACTTGAACAGCCCTGGAAGGCTTCACTTTTGAA 1418
Qy 1853 CTCTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTACAGCCACGATTCAGACGC 1912
Db 1419 CTCTTTCTGAGTGTAGATCTTTATATCCCCAGAACTTTACAGCAACAAATTCAGACTGC 1478
Qy 1913 TGCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGAGACATCTGTGTTCTGGCAGCTGTG 1972
Db 1479 TGCAGAAATTTGTTCTGAAGAGCCCTTCTGATTCAGAGAAGGATTTGTCTGCAAGTCTGA 1538
Qy 1973 TGTAGGGA---GCTGTGAGGAGGACCCCTTCCGGAGACCCCGACCATAGTTCCGGGGA 2029
Db 1539 TCTAGAGAAATTCCTCCAGCTGTGAAGCCTTCTGTGAGACCCCTGAGCATGTTCTGGGGA 1598
Qy 2030 GGAAGATGACTGGGAACCGAGTGCAGATGAAGCAGAGAATCTTAAATTTGTGGAATCTTT 2089
Db 1599 GGAAGATGACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAATCTGTGGAATCAT 1658
Qy 2090 CTGTCTATCTGAGAACCCCTTACAACTTTTAAATTTTAAAGCTCTCTTTTCAACCGTCAG 2149
Db 1659 CTGTAAATCTGTGATGACCCCTTACAACTTTTAAATTTTAAAGCTCTCTTTTCAAAATCAG 1718
Qy 2150 GAAGAAATTTGAAGGCGCTCAGGACTCAAGGCTCTTCTGAGTGCAGTGGCTCTCTC 2209
Db 1719 GGAATATGAGAAAGGCTGTGTGACTCAAGACCCCATCTGAGTGCATTTGCGCAATTC 1778
Qy 2210 TGGCCATCATACCTTACTTTCTGTAAGGCGCAGCTGTAGAGAGCAAGAATAATTG 2269
Db 1779 TGAGTGTCAACACCTTACTTTCTTGTAGTGTGAGTGTGGGAGCCAGAAAGTGAATG 1838
Qy 2270 TCCAGGCTGTGGGTGAGGCTCTTGTCTGGAGAAAGATACACCATATCAAGAA 2329
Db 1839 TCCAGACTCGGTACAGCGTGACGTTCTTCTGGAGGAAGACACACATGTCAAAAGAA 1898
Qy 2330 AAAGTAACTTCTCTGGAAGAGTTACTGTAGTATTATATAGTGGTGCATGAGGATCGCA 2389
Db 1899 AAAGTAACTTCTCTGGAAGAGTTACTGTAGTATTATATAGTGGTGCATGAGGATCGCA 1958
Qy 2390 AGGACCATGGGAAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 2449
Db 1959 AGGACCATGGGAAGAAATTTGCAAGGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 2018
Qy 2450 AGAAGTTGCCATTTGGCTTACTGTCTTACTTCTGAGCAGAGAAAGAAATTTTAAATAGACT 2509
Db 2019 AGAAGATGCTATTGGATATTGCTTGACATTTGGAACACAGAGAAAGAAATTTTAAATAGACT 2078
Qy 2510 ----GAGGATCGAGTCAAGGACTTACTGTGTGTACAGCAATGTTAAGAGTGA-ACAGCC 2564
Db 2079 CCAGGGAAATGCTTCAAGAGACTTAATGTTCTCAAGCAATGTTGAGTTGGCGCTGTA 2138
Qy 2565 TGCAACCGCGCCCACTCTGTCTTACTTCTGAGA-GTTTCCCTTAAAGAAACAAACACATGGC 2623
Db 2139 GTCTAGCTAGCATACACTACCTTCTTACCTGTAGAGGTGTCTTTTAAAGAAACAAATCTGGC 2198
Qy 2624 AGCTGTCTTGGACATGTTTTT 2645
```

```
Db 2199 AGCTGTCTTTTGACATTTTTTT 2220
RESULT 4
US-09-925-299-168
; Sequence 168, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 2324
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-168
Query Match 21.7%; Score 1189.2; DB 10; Length 2324;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 1642; Conservative 4; Mismatches 477; Indels 59; Gaps 11;
Qy 519 TTCGGGTGCGCTTCTTCTTGGCGATCGACGCTGTCTTTCGGAGTTCCGCGCGCTTCC 578
Db 43 TTTCAACCCCTCTCTCAGTGGCGATCGCAAGCAGGCTCTTCTAAGTTCCGAGCGCTTT 102
Qy 579 TCTCGACAAATCCCGGAACCTCGGCTCT-----GCCGAGCGTCGGACC 623
Db 103 GGCCCGGAAATCTCCGGAAACCCACACTGCTTTCTTGCCCGAGCCCGAGACTCGGGTC 162
Qy 624 AGGTACTTGAACAAATTTGCTTTCTCAGCTCTCTTGGCTGTCTCCTAGCTTATTCAGAAAG 683
Db 163 AGTTACTTGGAGAAATGCTCTCCAGCTCTTGGCGGCTCTCCCGGATTTGCTTCAGAA 222
Qy 684 CTGCTGCTTTGGAGCAGCTTTCCGGGGCTGATTTCTTACAGATGGCTAGATTTTGGC 743
Db 223 GTGCTAAATTTGGAGCAACTTTTTCGGTGGAAATTTTCCGACAGATGGCTAGATTTTGT 282
Qy 744 GCAAGTTTACAGCGCCCTGAGGCTTCGAGAGGAGGAGGAATCTGACGCTCCCAAGTG 803
Db 283 GGAGTCTACAGCGCCCTGAGAGCCCTTGAAGGGAGCGGAGAAACACAGCGCCCGCACAG 342
Qy 804 CAGAACTCTCTGAGTTTACACTGCGGCTGGACTCTTTCGCGAAGACTCGCGTCTGCTAGT 863
Db 343 CAGAAATCTTTGAGTT--CGCTGAGCTCGACTCTC---AGACCCCTCGGTCCAGTCC 398
Qy 864 CTTGCAATTTGGCTAGAGAGGAGCTCCAGTGGCAGTGTCTGTCTCAGA-CTGGAAGTTAA 922
Db 399 CTTGATTTGGCTAGAGAGGAGGATCCACTGGCAATACTGCGCCCGCAGACCTAAATAAT 458
Qy 923 ACTCAAGGCCCAGGAAGAGCTTTAGACTCTGACGCGCCCACTTTCTCTCTGGAGCAGCA 982
Db 459 GCTTAAAGGCCAAGGAAGTGTCTTTGACCCCTGAGCAGAGGCTTTTCTCTTAGAGCAGCA 518
Qy 983 GCTGTGGGAGTGGAGTGTGCTGCCAGTAGCCCTTCAAGCTGGTGTGCTAGTCTCTCCACCAG 1042
Db 519 GCTGTGGGAGTGGAGCTGTTCGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCGGA 578
Qy 1043 ACTTGACTTTTCAATCTCTTGGGCTCTGAGCGGTTTCAAGCTTAGGTAAATTTCAAGTAGT 1102
Db 579 ACTTGGCTCTTTCGCCCTCTGGGCTTCTAAACATTTCAACGATAGACAAATTTCAAGTGT 638
Qy 1103 TTCCTATCTCTGAACCCCTTCTTACCTGGACTTCTCCAGTTAGGCTGGCGTGTCA 1162
Db 639 ATCCTATTGCTGAACCCCTTCTTACCTGGACTGCTTTCTTAGGCTAGAAGTCAAGTATCA 698
```


QY 1959 TCTGCACTGTGTGTTAGGA---GCTGTGAGAGGAGACCCCTTCCGGAGACCCCGAC 2015
 DB 241 TCTGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTCTGAGACCCCTGAG 300
 QY 2016 CATAGTTCCGGGAGGAGATGACTGGGAAACCGAGTGCAGATGAAGCAGAGAAATCTTAAA 2075
 DB 301 CATAGTTCTGGGAGGAGATGACTGGGAAATCTAGTGCAGATGAAGCAGAGATCTTCAA 360
 QY 2076 ---TTGTGGAACTCTTCTGTCA---TTCTGAGGACCCCTCAACCTTTTAAATTTTAAAG 2129
 DB 361 ACTGTGGAACTCTTCTGTCTTAAATTCGTGATGACCCCTCAACCCCTTTTAAATTTTAAAG 420
 QY 2130 GCTCTTTTCAACCGTCAAGGAGAAATTTGGAAGCCGCTGAGACTCAAGGCTCTTCT 2189
 DB 421 GCTCTTTTCAACCATCAAGGAGAAATTTGGAAGGCTGTCTGAGACTCAAGGACCCCTATCT 480
 QY 2190 GAGGTCAAGTGGCTCTCTGGCCATCATACCTTTACTTTCTGTAAGGCCACCTGTTA 2249
 DB 481 GAGTCCATTGTGGCCATTTCTGAGTGTCAACCTTTACTTTCTGTAAGGTCGACTGTG 540
 QY 2250 GAGAGCCAAAGATAATTTGTCAGGCTGTGGGCTGGGTGAGGCTCTTGTCTGGAGAAAGA 2309
 DB 541 GGGAGCCAAAGATGTAATGTCTGAGTGTCCAGACTCGGTACAGGCTGACGTTCTTCTGGAGGAAGA 600
 QY 2310 TACACCCATCAAGAGAAAAGTAACTTCTCTGGAAGAGTTACTGAGTATTATATA 2359
 DB 601 CACACATGTCAAAAGAAAAGTAACTTCTCTGGAAGAGTTACTGAGTATTATATA 660
 QY 2370 AGTGTGATGAGGATCGCAAGGACCATGGGAAGATTTGCAAGGATGGATGAGGCTTC 2429
 DB 661 AGTGTGATGAGGATCGCAAGGACCATGGGAAGATTTGCAAGGATGGATGAGGCTTC 720
 QY 2430 CAGAAACGAATTCAGAAACAGAAAGTTGCCATTTGGCTACTCTTGGCTTTTGACACAGA 2489
 DB 721 CAGAAACGAATTCAGAAACAGAAAGTCTATTGATATCTTGGATTTGAACACAGA 780
 QY 2490 GAAAAATGTTTAAATAGACT---GAGATCGAGTCAAGGACTTACTGTTGTACAGCAA 2545
 DB 781 GAAAAATGTTTAAATAGACTCCAGGAAACATGCTTCAAGGACTTAATGTTCTCAAGCAA 840
 QY 2546 TGTTAAGAGTGAAC 2560
 DB 841 TGTGAGTTGCAGC 855

RESULT 6
 ; Sequence 3997, Application US/10756149
 ; Publication No. US20050181375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zlotnik, Albert
 ; APPLICANT: Aziiz, Natasha
 ; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
 ; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
 ; FILE REFERENCE: file
 ; CURRENT APPLICATION NUMBER: US/10756.149
 ; CURRENT FILING DATE: 2004-01-12
 ; NUMBER OF SEQ ID NOS: 5818
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 3997
 ; LENGTH: 465
 ; TYPE: DNA
 ; ORGANISM: Homo Sapiens
 ; US-10-756-149-3997

Query Match 3.9%; Score 215.8; DB 22; Length 465;
 Best Local Similarity 73.4%; Pred. No. 6e-48;
 Matches 343; Conservative 0; Mismatches 117; Indels 7; Gaps 5;

QY 758 CCTGAGAGCTTCGAGAGAGCGGAGGAATCTGACGCTCCACGCTGCAGAGAGTCTCTGAG 817
 DB 1 CCTGAGAGCCCTGAGGAGCGGAGAAACGAGCGCCGCCCAAGGAGAAATCTTTGAG 60

QY 818 TTACACTGCGGCTGAGACTCTTCCGGAAGAGACTCGGCTCGTCACTTCTGCAATTTGGCTAG 877
 DB 61 TT-CGCTGCACTCGACTCCTC---AGACCCCTCGGTCAACAGTCCCTTGTATTGGCTAG 116
 QY 878 AGAGGAGACTCCAGTGCAGTCTCGTCCCTCAGA-CTGGAAGTTTAAACTCAAGSCCCAGG 936
 DB 117 AGAGGAGATCCAGTGCAGTCTCGGCCCCAGACCTTAAATTTGGAGCTTTAAGGCCAAGG 176
 QY 937 AAAGAGCTTTAGACTCTGCAAGCGCCCACTTTCTCTCGGAGCAGCAGCTGTGGGGAGTGG 996
 DB 177 GAAGTGTCTTTGGACCCCTGCAAGCAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGGAGTGG 236
 QY 997 AGTTGCTGCCAGTAGAGCTTCAAGCTGTCTAGTCTCTCCACCGAGAACTTGCATCTTTCAT 1056
 DB 237 AGCTGTTGCCAGTAGAGCTTTCAATCCGCTGTACTCTAAACCGGGAACCTTGGGCTCTTCGC 296
 QY 1057 CCTCTGGGCTCTGAGCGTTTCAGAGCTTAGGTAATTTCAAGGTAGTTTCCCTATCTCTCTGA 1116
 DB 297 CCTCTGGGCTTTTAAACATTTCAAGCAGTAACTTTCAAGTGTGATATCTTATTGCTGA 356
 QY 1117 ACCCTTCTCTACCTGGACTACTTCCCGAGTTAGGCTGCGCTGTCTCAG-AGCAGCGCTGGA 1175
 DB 357 ACCCTTCTCTACCTGGACTGCTTTTCTAGGCTAGAGTCACTATCAGAAACAGTGTGGA 416
 QY 1176 GTGGGCCAGTTTGTGGGTTTCCGAAACAC-TAACCCCGCAGAGAGCTGCT 1221
 DB 417 AATAGCGAGGTAGTCTGGGCTTCCAGACACTTAAACCCCGAGAGAGCT 463

RESULT 7
 ; US-09-917-800A-1296/c
 ; Sequence 1296, Application US/09917800A
 ; Patent No. US20020119462A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mendrick, Donna
 ; APPLICANT: Porter, Mark
 ; APPLICANT: Johnson, Kory
 ; APPLICANT: Castle, Arthur
 ; APPLICANT: Blashoff, Michael
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5038-US
 ; CURRENT APPLICATION NUMBER: US/09/917, 800A
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR FILING DATE: 2000-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,040
 ; PRIOR FILING DATE: 2000-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,880
 ; PRIOR FILING DATE: 2000-11-02
 ; PRIOR APPLICATION NUMBER: US 60/290,029
 ; PRIOR FILING DATE: 2001-05-11
 ; PRIOR APPLICATION NUMBER: US 60/290,645
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: US 60/292,336
 ; PRIOR FILING DATE: 2001-05-22
 ; PRIOR APPLICATION NUMBER: US 60/295,798
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/297,457
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,884
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 60/303,459
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 1740
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1296
 ; LENGTH: 540
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI237580
 ; US-09-917-800A-1296

Query Match 3.3%; Score 181.6; DB 9; Length 540;

Query Match	2.2%;	Score 122;	DB 9;	Length 242;
Best Local Similarity	82.7%;	Pred. No. 2.3e-22;		
Matches 162;	Conservative 0;	Mismatches 32;	Indels 2;	Gaps 2;
QY	1982	CTGTCTAGAGAGGACCCCTTCCGGAGACCCCGGACCATAGTTCCGGGGAGGAGATGACTG	2041	

Db 22 CTCCAGTCTGGANGCCTTCTCTGAGACCCCTGAGCAGTAGTTCTGGGAGGAAGATGACTG 81
 Qy 2042 GGAACCCAGTGCAGATGAAGCAGAGAGTCTTAAATTTGGAACT-CTTTCTGTCTATCTG 2100
 Db 82 GGGATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGNACTCCATTTCTGTAACTTG 141
 Qy 2101 -AGAACCCCTACAACTTTTAAATTTTAAAGCTCTTTTCAACCGTCAGGGAAGATTCG 2159
 Db 142 AATGACCCCTACAACTTTTAAATTTTAAAGCTCTTTTCAAACTCAGGGAAGATGAG 201
 Qy 2160 AAAGCCGCTCAGGACT 2175
 Db 202 AAAGCTGTCTGTNT 217

RESULT 10

US-09-918-995-36363
 ; Sequence 36363, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR FILING DATE: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 36363
 ; LENGTH: 421
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-918-995-36363

Query Match 2.0%; Score 109; DB 10; Length 421;
 Best Local Similarity 71.6%; Pred. No. 1.3e-18;
 Matches 187; Conservative 0; Mismatches 65; Indels 9; Gaps 3;
 Qy 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTTATC--TAACCTTGA 5271
 Db 132 ACAGGTAAAGTTCCATTTCTGAGTGATGAATGTAACACTCTCTTCATCTTTAACTTGA 191
 Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325
 Db 192 ATCAAAACTATCAGATTTTATTTTGTATATTTTAAAGAGTAAAGTTAGGGACTAGA 251
 Qy 5326 ATTTTGTAGTTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAACTGGGAGATTGTAG 5384
 Db 252 AGACTCTAAATTTGGCTTCTACAGATCAATATTTTAAATGTAACCTAGTTGGGATTTATAG 311
 Qy 5385 CTCAGATATATTTGTATATATTTTAACTAACTCTGTAATTTGTAATAATATATTTGC 5444
 Db 312 ATAAATTTATTTTGGCGTATATAACATAACTAATCTGTAATTTGTAATAATATATTTGC 371
 Qy 5445 AATTATTAAAAA 5465
 Db 372 AATTATTAAAAA 392

RESULT 11

US-09-925-298-148
 ; Sequence 148, Application US/09925298
 ; Publication No. US20020039764A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: P103
 ; CURRENT APPLICATION NUMBER: US/09/925,298
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05881
 ; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 846
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 148
 ; LENGTH: 1028
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-298-148

Query Match 1.8%; Score 99.6; DB 9; Length 1028;
 Best Local Similarity 70.8%; Pred. No. 8.9e-16;
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;
 Qy 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTTATC--TAACCTTGA 5271
 Db 757 ACAGGTAAAGTTCCATTTCTGAGTGATGAATGTAACACTCTCTTCATCTTTAACTTGA 816
 Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325
 Db 817 ATCAAAACTATCAGATTTTATTTTGTATATTTTAAAGAGGTAAGTTAGGGACTAGA 876
 Qy 5326 ATTTTGTAGTTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAACTGGGAGATTGTAG 5384
 Db 877 AGACTCTAAATTTGGCTTCTACAGATCAATATTTTAAATGTAACCTAGTTGGGATTTATAG 936
 Qy 5385 CTCAGATATATTTGTATATATTTTAACTAACTCTGTAATTTGTAATAATATATTTGC 5444
 Db 937 TTAATTTATTTTGTATATTAACATACTAATCTGTAATTTGTAATAATATATTTGC 996
 Qy 5445 AATTATTAAA 5454
 Db 997 AATTATTAAA 1006

RESULT 12

US-10-102-806-148
 ; Sequence 148, Application US/10102806
 ; Publication No. US20030054421A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: P103P1C1
 ; CURRENT APPLICATION NUMBER: US/10/102,806
 ; CURRENT FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: 09/925,298
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05881
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 846
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 148
 ; LENGTH: 1028
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-102-806-148

Query Match 1.8%; Score 99.6; DB 14; Length 1028;
 Best Local Similarity 70.8%; Pred. No. 8.9e-16;
 Matches 177; Conservative 0; Mismatches 64; Indels 9; Gaps 3;
 Qy 5214 ACAGTGGGGCCACACTCTGAGGGCTGAAATGTGGCAACCCCTTTATC--TAACCTTGA 5271
 Db 757 ACAGGTAAAGTTCCATTTCTGAGTGATGAATGTAACACTCTCTTCATCTTTAACTTGA 816
 Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAAGAGTTGGGAATGAC 5325
 Db 817 ATCAAAACTATCAGATTTTATTTTGTATATTTTAAAGAGGTAAGTTAGGGACTAGA 876
 Qy 5326 ATTTTGTAGTTGGCCCTTTTCAGCTCAGTCATTT-TACGTGTAACTGGGAGATTGTAG 5384

Db 877 AGACTCTAAATGGCTCTACAGATCAATAAATTTAAATGTAACCTAGTTGGGATTTATAG 936
Qy 5385 CTCAGATTATATTTGTATATATTATTAACTAACTGTAATGCTGAAATGCTGTAATAATATATTTGC 5444
Db 937 TTAATAATATATTTGTGTATATATAACATACTAACTGTAATGCTGTAATGCTGTAATAATATATTTGC 996
Qy 5445 AATTATTAAA 5454
Db 997 AATTATTAAA 1006

RESULT 13

US-10-170-385-162
; Sequence 162, Application US/10170385
; Publication No. US2003020372A1
; GENERAL INFORMATION:
; APPLICANT: Mundy, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 53268200100
; CURRENT APPLICATION NUMBER: US/10/170,385
; PRIORITY FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-170-385-162

Query Match 1.8%; Score 98; DB 17; Length 2148;

Best Local Similarity 70.4%; Pred. No. 4.1e-15;

Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;

Qy 5214 ACAGTGGGGCCACACTTCTGAGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271
Db 1888 ACAGTAAAGTTCCATTTCTGAGTGATGAAATGTAACACTTCTTTCATCTTTAACTTGAA 1947
Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAGAGTTGGGGAATGAC 5325
Db 1948 ATCAAAACTATCAGATTTTATTTTGTATATTTAAGGAAGGTAAGTTAGGGGACTAGA 2007
Qy 5326 ATTTTTCAGTTGGCCCTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGATAG 5384
Db 2008 AGACTCTAAATGGCTCTACAGATCAATAATTTAAATGTAACCTAGTTGGGATTTATAG 2067
Qy 5385 CTCAGATTATTTGTATATATAATTAATTAACACTGTAATCTGTAATAATATATATTTGC 5444
Db 2068 TTAATAATATATTTGTGTATATAACATAATTAATCTGTAATCTGTAATAATATATATTTGC 2127
Qy 5445 AATTATTAAA 5454
Db 2128 AATTATTAAA 2137

RESULT 14

US-10-131-827-8849
; Sequence 8849, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay

; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; PRIORITY FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8849
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8849

Query Match 1.8%; Score 98; DB 17; Length 2148;

Best Local Similarity 70.4%; Pred. No. 4.1e-15;

Matches 176; Conservative 0; Mismatches 65; Indels 9; Gaps 3;

Qy 5214 ACAGTGGGGCCACACTTCTGAGGCTGAAATGTGGCAACCCCTTATC--TAACTTGAA 5271
Db 1888 ACAGTAAAGTTCCATTTCTGAGTGATGAAATGTAACACTTCTTTCATCTTTAACTTGAA 1947
Qy 5272 ATCAAAACCGTCAAAATTTTATTTT-----TATAATTTAAGAAGAGTTGGGGAATGAC 5325
Db 1948 ATCAAAACTATCAGATTTTATTTTGTATATTTAAGGAAGGTAAGTTAGGGGACTAGA 2007
Qy 5326 ATTTTTCAGTTGGCCCTTTCAGCTCAGTCATTT--TACGTGTAACGTGGAGATTGATAG 5384
Db 2008 AGACTCTAAATGGCTCTACAGATCAATAATTTAAATGTAACCTAGTTGGGATTTATAG 2067
Qy 5385 CTCAGATTATTTGTATATATAATTAATTAACACTGTAATCTGTAATAATATATATTTGC 5444
Db 2068 TTAATAATATATTTGTGTATATAACATAATTAATCTGTAATCTGTAATAATATATATTTGC 2127
Qy 5445 AATTATTAAA 5454
Db 2128 AATTATTAAA 2137

RESULT 15

US-09-917-800A-1399
; Sequence 1399, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; PRIORITY FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457

; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1399
; LENGTH: 8351
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 M11794
US-09-917-800A-1399

Query Match	1.5t;	Score 79.4;	DB 9;	Length 8351;
Best Local Similarity	64.8t;	Pred. No. 1.3e-09;		
Matches 138;	Conservative 0;	Mismatches 66;	Indels 9;	Gaps 1;

Qy	3504	GTATTGGTTTGTGTTGTTGTTTATTTTACATTTTCATAGATTTGACTTAATGAAGGCAA	3563
Db	6264	GTATTGTTTTTCTTTTTTTTGTGTTTATTTGTTTGTGTTTATCTAATGCAG	6323
Qy	3564	AAACCTGTATCAACCTAAAGACATTGATGTCTCACTTCAGTGGTGGATTTCTCCCTC	3623
Db	6324	TCCCTGGATATCACCTAAATGATCCCTCTGCCTCGGTTTTTTTTTTTTTTTTTTT	6383
Qy	3624	TTTTTTTTTTTCCCC-----CAGAGCTGAGGACCAACCCAGGCTTTGCACTTGC	3674
Db	6384	TTTTTTTTTTTGGTTCTTTTTTCGGAGCTGGGACCGAACCCAGGCGCTTGGCGCTCC	6443
Qy	3675	TAGCAAGCGCTCTACCACTGAGCTAAATCCCC	3707
Db	6444	TAGTAAGCGCTCTACCACTGAGCTAAATCCTC	6476

Search completed: September 16, 2005, 10:40:03
Job time : 3240.64 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 11:51:14 ; Search time 16323.4 Seconds
(without alignments)
12750.757 Million cell updates/sec

Title: US-10-650-482-3
Perfect score: 5468
Sequence: 1 cggctccgcgtcgcctgcg.....attaaaaaaaaaaaaaaa 5468

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_ges1: *
9: gb_ges2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2641	48.3	2731	3	AK049028 Mus muscu
2	2522.8	46.1	2619	3	AK082957 Mus muscu
3	1271	23.2	1342	3	AK086606 Mus muscu
4	828.2	15.1	871	5	BQ922494 AGENCOURT
5	810	14.8	966	5	BQ926670 AGENCOURT
6	808.6	14.8	997	1	AV055059 AV055059
7	799.2	14.6	890	7	CF583619 AGENCOURT
8	796	14.6	1011	2	BB610365 BB610365
9	794.2	14.5	913	7	CF583620 AGENCOURT
10	786.6	14.4	945	6	CA978373 AGENCOURT
11	786.4	14.4	3941	3	HSM805059
12	774.8	14.2	801	7	CO428800 UI-M-HX0
13	767.4	14.0	796	4	BI654191 603280489
14	763	14.0	854	7	CF746885 UI-M-HX0
15	759.6	13.9	835	7	CF617044 AGENCOURT
16	754.8	13.8	936	4	BI737283 603357931
17	739.8	13.5	876	5	BQ953605 AGENCOURT
18	737.4	13.5	954	5	BQ943218 AGENCOURT
19	732.8	13.4	811	6	CA750998 UI-M-HX0
20	730	13.4	744	7	CF749066 UI-M-HX0
21	720.4	13.2	725	6	CD352623 UI-M-HX0
22	717.2	13.1	784	7	CK021950 AGENCOURT
23	717	13.1	717	5	BQ444408 UI-M-HX0
24	715.8	13.1	731	7	CO426562 UI-M-HX0

CF732867	UI-M-HB0-
BU057555	UI-M-FO0-
CF583507	AGENCOURT
CF736709	UI-M-HD0-
BF608942	MY1 00188
BB527493	BB527493
CA456136	UI-M-HN0-
CF538316	UI-M-G10-
BQ897619	AGENCOURT
CF616403	AGENCOURT
CA457283	UI-M-HN0-
CF745488	UI-M-GV0-
CA460404	UI-M-HB0-
CB575918	AGENCOURT
CN530841	UI-M-HQ0-
CF531425	UI-M-FV0-
CD348078	UI-M-FV0-
CA318387	UI-M-FW0-
CF737779	UI-M-HD0-
CF537499	UI-M-G10-
CO430576	UI-M-HD0-
CF729004	UI-M-HX0-
CB594340	AGENCOURT
CN698894	E0414804
CG974327	E02844078
BI080754	602878517
CD808262	UI-M-GW0-
CN527875	UI-M-HQ0-
CN721807	E0830A09-
AV297289	AV297289
BE648121	UI-M-BH2-
BI691569	603307237
CN522421	UI-M-HQ0-
BB395632	BB395632
BQ570910	UI-M-FB0-
BF780757	602103031
BU058221	UI-M-FO0-
BI104915	602892704
CF739087	UI-M-HD0-
CN699271	E0418B02-
BG695164	NISC lv13
BI648124	603278442
BI554632	603235979
BI651975	603298930
BI081605	602880239
CN701167	E0445H07-
CF537328	UI-M-G10-
BB651558	BB651558
BF539497	602049672
BY757855	BY757855
BY726451	BY726451
BY737535	BY737535
CK129449	AGENCOURT
BY757590	BY757590
CB208832	AGENCOURT
BQ605339	AGENCOURT
BI696902	603348641
BM249564	K0835E02-
CN698796	E0412H08-
CN698868	E0413B07-
BY737242	BY737242
BF099286	601752463
BB663346	BB663346
CN666892	A0846D03-
BI153462	602916365
BF178507	601807668
BQ567519	G191911.Y
BE283428	601103424
BP764101	BP764101
AW543985	C0176B05-
BZ090269	CH230-233
CF898558	A0441A04-
CK627587	I005H12.Y

98	575.4	10.5	610	6	CD560629	B0416D08-BE914017 601665429
99	575.2	10.5	1079	2	BE914017	601665429
100	575	10.5	575	4	B1713435	ic88e10.Y

RESULT 1	
LOCUS	AK049028
DEFINITION	Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length enriched library, clone:C230093D15 product:hypothetical protein, full insert sequence.
ACCESSION	AK049028
VERSION	AK049028.1 GI:26339691
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, F., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2731)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, R., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission

JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/.
FEATURES	source 1. .2731 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM DB:C230093D15" /db_xref="taxon:10090" /clone="C230093D15" /tissue_type="cerebellum" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="0 day neonate" 286..2379 /notes="unnamed protein product; hypothetical protein (evidence: ProCrest)" /codon_start=1 /protein_id="BAC33517.1" /db_xref="GI:26339692" /translation="MTGTGTHRRKRPGLRSGWFLPFLRRSHACSEFPFPPSRQNP GNSALPERTRYWTKLSLLQLALPLSFQKLLWSQLFGGLIPTRWLDFAASVALRA LRGSAAPFTQKLSLSDSSDLVWSLDWLEGLQVSSDLEKFKLQAEQA LDAAAPFTLLQGLWGLVPSLLQAGLVSHRELDSSGLSPVSLNFKVSYLLN PSYLDVLPGLKQSSAGGQFVGFRTLTPTSCYLSGDSCHPQPLRAEASVATWRR PPLSTGLPEIHLRMKRLFLQANKGOELPTDQNGYHSLREHNLRLMDPOHCTD NPAQVSPADRPETPEKPELVQVVSQPSQSSLPCELPVEKECEDHTNATDSD RGSLSPVTPVCSNKLIDYILGAFSDLEASDSESDWGERPEDGPDGSDGLSSQ DVQDSEGLHWNFSHVDYNPQNTATTAARIAPRDPDSGTSWSGSGVSGSQ EGLPPTPDHSGEEDDWEPSADEANLKNWFCHESEDYFNILNFKAPFPQSGKNWK GRODSKASSEATVAFSGHHTLLSKAQLLESQEDNCPGCGLGALAGERVTHIKRKV TFLEVTVEYVYISGDEDRKGPWFERFARDGCRFKRIQTEVAIGVCLAFEHREKMFNL RIESKDLLLYSNVKK"
CDS	

ORIGIN	Query Match 48.3%; Score 2641; DB 3; Length 2731; Best Local Similarity 99.1%; Pred. No. 0; Matches 2698; Conservative 0; Mismatches 20; Indels 5; Gaps 4;
Qy	178 TCTCTATGTTGGAGCGATCTCACAGCGCTAGGACGTCTCTTCCCTAGCGGGATGGAC 237
Db	2 TCTCTATGTTGGAGCGATCTCACAGCGCTAGGACGTCTCTTCCCTAGCGGGATGGAC 61
Qy	238 CTAACCGCGTCCGCCCGCTTGGCGGGCTCTTGGCGGCTCGGTCGAGCAGCTCGTTGC 297
Db	62 CTAACCGCGTCCGCCCGCTTGGCGGGCTCTTGGCGGCTCGGTCGAGCAGCTCGTTGC 121
Qy	298 GGAAGCGCGCGCTCTCTTGGGCTCTCTTGGCGGCGGGAAATCGGACTGACGATCCCACT 357
Db	122 GGAAGCGCGCGCTCTCTTGGGCTCTCTTGGCGGCGGGAAATCGGACTGACGATCCCACT 181
Qy	358 CGGTGGTGGGCAAGCGGAGACTGTGTAGACCTCGGATCGAGCTCGGCTGAGCGCGCT 417
Db	182 CCGTGGTGGGCAAGCGGAGACTGTGTAGACCTCGGATCGGCTCGGCTGAGCGCGCT 241
Qy	418 GAGCTCTGTCTCTCTCTGTCTGAGAGCGCCCAAGAAAGGAGATGAGACAGAAACGC 477
Db	242 GAGCTCTGTCTCTCTCTGTCTGAGAGCGCCCAAGAAAGGAGATGAGACAGAAACGC 301
Qy	478 ACAGGGCCCGGAAAGCGGCTTGGGCTCTCTGTGGGCTCTCTGTGGGCTCTCTTCTTCTTC 537

|||||
302 ACAGGGCCGGAAGGGCTGGCCCTCGGCTGGGCTCTGGTTCCGGCTGCCCTTCCTTC 361
QY
538 GGGCATCGCAGCGCTGCTCTTCGGAGTTCGCGCGCTTCTCTCGACAAATCCCGGA 597
Db
362 GGGCATCGCAGCGCTGCTCTTCGGAGTTCGCGCGCTTCTCTCGACAAATCCCGGA 421
QY
598 ACTCCGCTCTGCCGAGCGTCGACACAGGTACTGGACCAATTCCTTCTCAGCTCCTTG 657
Db
422 ACTCCGCTCTGCCGAGCGTCGACACAGGTACTGGACCAATTCCTTCTCAGCTCCTTG 481
QY
658 CCCTGCTCCCTAGCCTATTCAGAAGCTGTGCTTTGGAGCCAGCTTTCGGGGGCTGA 717
Db
482 CCCTGCTCCCTAGCCTATTCAGAAGCTGTGCTTTGGAGCCAGCTTTCGGGGGCTGA 541
QY
718 TTCTTACCAAGTGTAGATTTTCGCCCAAGTTACAGCGCCCTCAGAGCTTCGAGAGGAC 777
Db
542 TTCTTACCAAGTGTAGATTTTCGCCCAAGTTACAGCGCCCTCAGAGCTTCGAGAGGAC 601
QY
778 GGGAGGAATCTGACGCTCCCAAGGTGAGAGTCTCTGAGTTACACTGCGGCTGGACTCT 837
Db
602 GGGAGGAATCTGACGCTCCCAAGGTGAGAGTCTCTGAGTT-CACTGCGGCTGGACTCT 660
QY
838 TCGGAAGACTT- -GGTTGTCAAGTCTCTTG-ATTGGCTTAGAGGAGGACTGCAGTGGCAG 897
Db
661 TCGGAAGACTT- -GGTTGTCAAGTCTCTTG-ATTGGCTTAGAGGAGGACTGCAGTGGCAG 717
QY
898 TGCTGCTCTCAGACTCGAGT- -AACTCAAGGCCAGGAAGAGCTTTAGACTCTGCA 956
Db
718 TGCTGCTCTCAGACTCGAGT- -AACTCAAGGCCAGGAAGAGCTTTAGACTCTGCA 777
QY
957 GCGCCCACTTTCTCTCGGAGCAGCTGTGGGGAGTGGAGTTGCTGCCAGTAGCCTT 1016
Db
778 GCGCCCACTTTCTCTCGGAGCAGCTGTGGGGAGTGGAGTTGCTGCCAGTAGCCTT 837
QY
1017 CAAGCTGTCTAGTCTCCCAAGAGACTTGAATCTTCACTCTCGGAGTTCAGAGCTT 1076
Db
838 CAAGCTGTCTAGTCTCCCAAGAGACTTGAATCTTCACTCTCGGAGTTCAGAGCTT 897
QY
1077 CAGAGCTTAGGTAAATTCAGGTAGTTTCTATCTCTGAACTCTCTCTCTCTCTCTCTCT 1136
Db
898 CAGAGCTTAGGTAAATTCAGGTAGTTTCTATCTCTGAACTCTCTCTCTCTCTCTCTCT 957
QY
1137 CTTCGCCAGTTAGGCTGCGCTGTGAGAGCAGCTGTGAGTGGCCAGTTTGTGGTTTC 1196
Db
958 CTTCGCCAGTTAGGCTGCGCTGTGAGAGCAGCTGTGAGTGGCCAGTTTGTGGTTTC 1017
QY
1197 CGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAAGTGGTTGCTCAGCCGCTTG 1256
Db
1018 CGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAAGTGGTTGCTCAGCCGCTTG 1077
QY
1257 CGGCGAGAGATGTGGCAACCGCTGGAGAGGTGTCCGCTCTCTCTACAGAGGCTTG 1316
Db
1078 CGGCGAGAGATGTGGCAACCGCTGGAGAGGTGTCCGCTCTCTCTCTACAGAGGCTTG 1137
QY
1317 CCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTCCTCCAGGCTTAAACAGGGCAA 1376
Db
1138 CCGGAAATCCACCACTTCGTATGAAACGGCTAGAAATTCCTCCAGGCTTAAACAGGGCAA 1197
QY
1377 GAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAACATAACCTT 1436
Db
1198 GAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTTGGAGGAGGAACATAACCTT 1257
QY
1437 CTCGGATGGACCCACACATTCACAGATTAACCCAGCAGCGGCTGCTCCCTGCTGCA 1496
Db
1258 CTCGGATGGACCCACACATTCACAGATTAACCCAGCAGCGGCTGCTCCCTGCTGCA 1317
QY
1497 GACAGGCGGAGCCCACTGAGAAAAAACAGAAATTTGGTATTCAGAAAGTTTCAAGAGC 1556
Db
1318 GACAGGCGGAGCCCACTGAGAAAAAACAGAAATTTGGTATTCAGAAAGTTTCAAGAGC 1377
QY
1557 CCCAGGGAAGCAGTCTGTTTTGTGAATTAACCGTGGAAAAAGAAATGTGAAGAGGACAC 1616
|||||

Db
1378 CCCAGGGAAGCAGTCTGTTTTGTGAATTAACCGTGGAAAAAGAAATGTGAAGAGGACCAC 1437
QY
1617 ACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTCTGTTTCTACAGACCAAGTT 1676
Db
1438 ACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTCTGTTTCTACAGACCAAGTT 1497
QY
1677 TGTAAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCCCAGTGTGATTTGGAAGCAGC 1736
Db
1498 TGTAAGCAACAAACTGATAGATTATATTTTGGGAGGCGCCCCAGTGTGATTTGGAAGCAGC 1557
QY
1737 TCTGATTTCTGAAGTGGGATTCGGGCGAGGAACCTGAGACGATGGCTTTGTATAGCGAT 1796
Db
1558 TCTGATTTCTGAAGTGGGATTCGGGCGAGGAACCTGAGACGATGGCTTTGTATAGCGAT 1617
QY
1797 GGGTCCCTGTCTGAATCAGACGTCGGAACAGGAATCGGAAGGCTTCACTTTTGGAACTCT 1856
Db
1618 GGGTCCCTGTCTGAATCAGACGTCGGAACAGGAATCGGAAGGCTTCACTTTTGGAACTCT 1677
QY
1857 TTCCAAGTGTAGATCTCTTACAAACCCCAAAATTTTACAGCCAGATTCAGACGGTGGC 1916
Db
1678 TTCCAAGTGTAGATCTCTTACAAACCCCAAAATTTTACAGCCAGATTCAGACGGTGGC 1737
QY
1917 AGAATTCGCCCCCAGAGACCCATCAGATTCAGGACATCTCTGCTGTCGACGCTGTGGTGA 1976
Db
1738 AGAATTCGCCCCCAGAGACCCATCAGATTCAGGACATCTCTGCTGTCGACGCTGTGGTGA 1797
QY
1977 GGGAGCTGTGAGAGGAGCCCTTCCGGAGACCCCGACCATAGTTTCCGGGAGGAAGAT 2036
Db
1798 GGGAGCTGTGAGAGGAGCCCTTCCGGAGACCCCGACCATAGTTTCCGGGAGGAAGAT 1857
QY
2037 GACTGGGAACCCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAACTCTCTTCTGTCT 2096
Db
1858 GACTGGGAACCCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGTGGAACTCTCTTCTGTCT 1917
QY
2097 TCTGAGAGCCCTTACAACTTTTAAATTTTAAAGCTCTCTTCAACCGTCAGGGAAGAT 2156
Db
1918 TCTGAGAGCCCTTACAACTTTTAAATTTTAAAGCTCTCTTCAACCGTCAGGGAAGAT 1977
QY
2157 TGGAAAGGCCGTCAAGGACTCAAGGCTCTTCTGAGGTCAAGTGGCTTCTCTGCGCAT 2216
Db
1978 TGGAAAGGCCGTCAAGGACTCAAGGCTCTTCTGAGGTCAAGTGGCTTCTCTGCGCAT 2037
QY
2217 CATACCTTACTTTCTTGTAAAGCCAGCTGTGTAGAGCCAAAGAAATATTTGTCCAGGC 2276
Db
2038 CATACCTTACTTTCTTGTAAAGCCAGCTGTGTAGAGCCAAAGAAATATTTGTCCAGGC 2097
QY
2277 TGTGGGCTGGGTGAGGCTCTTGTGGAGAAAGATAACCCATATCAAGAGAAAAAGGTA 2336
Db
2098 TGTGGGCTGGGTGAGGCTCTTGTGGAGAAAGATAACCCATATCAAGAGAAAAAGGTA 2157
QY
2337 ACCTTCTCGGAAGAGTTACTGAGTATATATTAAGTGGTGTATGAGGATCGCAAGGACCA 2396
Db
2158 ACCTTCTCGAGAGAGTTACTGAGTATATTAAGTGGTGTATGAGGATCGCAAGGACCA 2217
QY
2397 TGGGAAGAAATTTGCAAGGAGTGTATGAGGTTTCAGAAAAAGAAATTCAGAAAAACAGAGTT 2456
Db
2218 TGGGAAGAAATTTGCAAGGAGTGTATGAGGTTTCAGAAAAAGAAATTCAGAAAAACAGAGTT 2277
QY
2457 GCCATTTGGCTTACTTGTGGCTTTGAGCAGAGAAAAAATGTTTATAGACTGAGGATC 2516
Db
2278 GCCATTTGGCTTACTTGTGGCTTTGAGCAGAGAAAAAATGTTTATAGACTGAGGATC 2337
QY
2517 GAGTCAAGGACTTACTTGTGTACAGCAATGTTTAAAGTGAAGTGAACAGGCTCGAACCCGTGC 2576
Db
2338 GAGTCAAGGACTTACTTGTGTACAGCAATGTTTAAAGTGAAGTGAACAGGCTCGAACCCGTGC 2397
QY
2577 CCACTCTGCTCTTACTTGTAGAGTTTCCCTTAAAAACAAACACTGGCAGCTGTCTTGGGA 2636
Db
2398 CCACTCTGCTCTTACTTGTAGAGTTTCCCTTAAAAACAAACACTGGCAGCTGTCTTGGGA 2457
QY
2637 CATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAATTTGATTTATTTTGGGTATGT 2696
Db
2458 CATGTTTTTAAAGAAACAACTTGTATCTAGAGATGCAATTTGATTTATTTTGGGTATGT 2517
|||||

Qy	226	GC	CGGGATGGACCTTAACCGCGGTCCGCAACCGCTTTCGCGGCGCTCTGGCGGTC	285
Db	182	GC	CGGGATGGACCTTAAACCGCGGTCCGCAACCGCTTTCGCGGCGCTCTGGCGGTC	241
Qy	286	AG	CACCTGTTGGGAAGCGCGCTCTCTGGGCTCTCTGCGCGCGGGAATCGGACT	345
Db	242	AG	CACCTGTTGGGAAGCGCGCTCTCTGGGCTCTCTGCGCGCGGGAATCGGACT	301
Qy	346	GC	GTACCCACTCCGCTGGCGAAGCGGAGACTGTGTAGACTCTGGATCCAGCCTGC	405
Db	302	GC	GTACCCACTCCGCTGGCGAAGCGGAGACTGTGTAGACTCTGGATCCAGCCTGC	361
Qy	406	GCT	GACGCGCTGAGCTCTGTCTCTCTCTGTGTGAGAGCGCGCAAGGAAGAGATGG	465
Db	362	GCT	GACGCGCTGAGCTCTGTCTCTCTCTGTGTGAGAGCGCGCAAGGAAGAGATGG	421
Qy	466	AG	CAGGAAACGACAGGCGCGGAGCGGCTGCGCTCGGCTGGGCTCTGTTCCGGC	525
Db	422	AG	CAGGAAACGACAGGCGCGGAGCGGCTGCGCTCGGCTGGGCTCTGTTCCGGC	481
Qy	526	TG	CCCTTCTTCCGCGATCGCACGCTGCTCTCGGAGTTCCCGCGCTTCTCTCTCGAC	585
Db	482	TG	CCCTTCTTCCGCGATCGCACGCTGCTCTCGGAGTTCCCGCGCTTCTCTCTCGAC	541
Qy	586	AAAA	TCCCGGAACTCCGCTCTGCGCGAGCGTCCGACCAAGTACTGGACCAAAATTCGTTT	645
Db	542	AAAA	TCCCGGAACTCCGCTCTGCGCGAGCGTCCGACCAAGTACTGGACCAAAATTCGTTT	601
Qy	646	CT	CAGCTCTTCCCTGCTCCTAGCCTATTCAGAGAGCTGCTCTTGGAGCAGCTTT	705
Db	602	CT	CAGCTCTTCCCTGCTCCTAGCCTATTCAGAGAGCTGCTCTTGGAGCAGCTTT	661
Qy	706	CC	GGGCGCTGATTTCTACAGATGGCTAGATTTTGGCGAAGTTACAGCGCCCTGAGAG	765
Db	662	TC	GGGCGCTGATTTCTACAGATGGCTAGATTTTGGCGAAGTTACAGCGCCCTGAGAG	721
Qy	766	CT	TCAGAGGACGGGAGAACTTCAACGCTCCACGGTGCAGAGTCTCTGAGTTTACATG	825
Db	722	CT	TTGAGAGGACGGGAGAACTTCCGCTCCACGGTGCAGAGTCTCTGAGTTTACATG	780
Qy	826	CG	CTGAGCTCTTCCGAGAGACTCGGCTGCTCAGTACTCTGCAATTCGCTAGAGAGGA	885
Db	781	CG	CTGAGCTCTTCCGAGAGACTTGGTGTGCTCTCTTTG - ATTGGCTAGAGAGGGA	837
Qy	886	CT	CAGTGGCAGTCTGCTCTCAGA - CTGGAAGTTAAACTCAAGGCCACAGAAAGACT	944
Db	838	CT	CAGTGGCAGTCTGCTCTCAGA - CTGGAAGTTAAACTCAAGGCCACAGAAAGACT	897
Qy	945	TT	AGACTCTGACGCGCCACTTTCTCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTG	1004
Db	898	TT	AGACTCTGACGCGCCACTTTCTCTGAGCAGCAGCTGTGGGAGTGGAGTTGCTG	957
Qy	1005	CC	CAGTGGCTTCAAGCTGCTAGTCTCCACCGAGAACTTGACTCTCATCTCTGGG	1064
Db	958	CC	CAGTGGCTTCAAGCTGCTAGTCTCCACCGAGAACTTGACTCTCATCTCTGGG	1017
Qy	1065	CCT	CTGAGCTTCAAGCTTAAATTTCAAGTGTGTTCTCTCTCTGAAACCTTCC	1124
Db	1018	CCT	CTGAGCTTCAAGCTTAAATTTCAAGTGTGTTCTCTCTCTGAAACCTTCC	1077
Qy	1125	TAC	TGGACTACTTCCCGCTTGGGCTGGCTGTGAGCAGCGCTGGAGGTGGCCAG	1184
Db	1078	TAC	TGGACTACTTCCCGCTTGGGCTGGCTGTGAGCAGCGCTGGAGGTGGCCAG	1137
Qy	1185	TT	TGGGTTTCCGAACTAAACCCAGAGAGCTGCTATCTTTCTGAAGATGGTGTGAC	1244
Db	1138	TT	TGGGTTTCCGAACTAAACCCAGAGAGCTGCTATCTTTCTGAAGATGGTGTGAC	1197
Qy	1245	CCT	CAGCGTTGGCGGAGAGATGCGCAACCGCTGGGAGAGTGTCCGCTCTCTCT	1304
Db	1198	CCT	CAGCGTTGGCGGAGAGATGCGCAACCGCTGGGAGAGTGTCCGCTCTCTCTCT	1257

Qy	1305	AC	AGAGGCTTCGCGGAATCCACCACTTCTGTATGAACCGCTAGAAATTCCTCCAGGCT	1364
Db	1258	AC	AGAGGCTTCGCGGAATCCACCACTTCTGTATGAACCGCTAGAAATTCCTCCAGGCT	1317
Qy	1365	AC	AAAGGGCAAGTGTACCCACCTGACCAAGATAATGGCTATCATAGCTTGGAGGAG	1424
Db	1318	AC	AAAGGGCAAGTGTACCCACCTGACCAAGATAATGGCTATCATAGCTTGGAGGAG	1377
Qy	1425	GA	ACATAACCTTCTCCGGATGGACCAACATTTGACAGATAACCCAGACACAGGGGTG	1484
Db	1378	GA	ACATAACCTTCTCCGGATGGACCAACATTTGACAGATAACCCAGACACAGGGGTG	1437
Qy	1485	TC	CCCTGCTGCAGACAGGCGGAGCCCACTGAGAAAACCAAAATTTGGTATTCAAGAA	1544
Db	1438	TC	CCCTGCTGCAGACAGGCGGAGCCCACTGAGAAAACCAAAATTTGGTATTCAAGAA	1497
Qy	1545	GT	TTTCAGAGCCCCCAGAGGAGCAGTCTGTTTTGTGAATTTACCCGTGGAAAAAGATGT	1604
Db	1498	GT	TTTCAGAGCCCCCAGAGGAGCAGTCTGTTTTGTGAATTTACCCGTGGAAAAAGATGT	1557
Qy	1605	GA	AGAGACCACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTGTTTCT	1664
Db	1558	GA	AGAGAGCACACTAATGCAACTGACCTCTCAGATAGAGGAGAGCCTTCTGTTTCT	1617
Qy	1665	AC	CAGACCTTTGTAGCAAACTGATAGATTATTTTGGAGGCGCCCCCAGTGAC	1724
Db	1618	AC	CAGACCTTTGTAGCAAACTGATAGATTATTTTGGAGGCGCCCCCAGTGAC	1677
Qy	1725	TT	CGAGCCAGCTCATTTCTGAAAGTGGATTTGGGGGAGGAACTCTAGGAGCATGGC	1784
Db	1678	TT	CGAGCCAGCTCATTTCTGAAAGTGGATTTGGGGGAGGAACTCTAGGAGCATGGC	1737
Qy	1785	TT	TGATAGCAGTGGCTCCCTGCTGCACTGACCTCTCAGATAGAGGAGAGCCTTCTGTTTCT	1844
Db	1738	TT	TGATAGCAGTGGCTCCCTGCTGCACTGACCTCTCAGATAGAGGAGAGCCTTCTGTTTCT	1797
Qy	1845	CT	TTGGAACCTTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTACAGCCAGATT	1904
Db	1798	CT	TTGGAACCTTTTCCACAGTGTAGATCTTTACAAACCCCAAACTTTACAGCCAGATT	1857
Qy	1905	CAG	ACGCTCCCAAGATTGCCCCCAGAGCCCATCAGATTCAGGAGCATCTGCTCTGGC	1964
Db	1858	CAG	ACGCTCCCAAGATTGCCCCCAGAGCCCATCAGATTCAGGAGCATCTGCTCTGGC	1917
Qy	1965	AG	CTGTGTGTAGGAGCTGTCTAGGAGGAGCCCTTCCGAGACCCCGCCAGCATAGTTC	2024
Db	1918	AG	CTGTGTGTAGGAGCTGTCTAGGAGGAGCCCTTCCGAGACCCCGCCAGCATAGTTC	1977
Qy	2025	GG	GAGAGATGACTGGGAAACCGAGTGCAGATGAAGCAGAGAACTTTAAATTTGTGGAAC	2084
Db	1978	GG	GAGAGATGACTGGGAAACCGAGTGCAGATGAAGCAGAGAACTTTAAATTTGTGGAAC	2037
Qy	2085	TC	TTTCTGCTCATTCAGGAGCCCTTCAACCTTTTAAATTTTAAAGCTCTCTTTTCAACCG	2144
Db	2038	TC	TTTCTGCTCATTCAGGAGCCCTTCAACCTTTTAAATTTTAAAGCTCTCTTTTCAACCG	2097
Qy	2145	TC	AGGAAGAAATTTGGAAGGCGCTCAGGACTCAAGAGCCCTTCTGAGGTCACAGTGGCC	2204
Db	2098	TC	AGGAAGAAATTTGGAAGGCGCTCAGGACTCAAGAGCCCTTCTGAGGTCACAGTGGCC	2157
Qy	2205	TT	CTCTGGCCCATCATACCTTCTTTGTAAGCCAGCTGTGTAGAGCCCAAGAGAT	2264
Db	2158	TT	CTCTGGCCCATCATACCTTCTTTGTAAGCCAGCTGTGTAGAGCCCAAGAGAT	2217
Qy	2265	AA	TTGTCAGGCTGGGCTGGGCTGAGGCTTTGCTGGGAAAGATACACCCATATCAAG	2324
Db	2218	AA	TTGTCAGGCTGGGCTGGGCTGAGGCTTTGCTGGGAAAGATACACCCATATCAAG	2277
Qy	2325	AG	AAAAAGTAACCTTCTCTGGAAGAGTGTACTGAGTATTATATAAGTGTGTGAGGAT	2384
Db	2278	AG	AAAAAGTAACCTTCTCTGGAAGAGTGTACTGAGTATTATATAAGTGTGTGAGGAT	2337
Qy	2385	CG	AAAGGACCATGGGAGAAATTTTGAAAGGAGTGTGAGGTTCCAGAAACGAAATTCAA	2444

```

|||||
2338 CGCAAGGACCATGGGAAGATTTTCCAAAGGATGATGCAGGTTCCAGAAACGAATTCAA 2397
QY 2445 GAAACAGAGTGGCCATTGGCTACTGCTTGGCCCTTTGAGCAGACAGAAAAATGTTTAAT 2504
Db 2398 GAAACAGAGTGGCCATTGGCTACTGCTTGGCCCTTTGAGCAGACAGAAAAATGTTTAAT 2457
QY 2505 AGACTGAGGATCGAGTCAAGGACTTACTGTGTGTACAGCAATGTTTAAAGAGTGAACAGCC 2564
Db 2458 AGACTGAGGATCGAGTCAAGGACTTACTGTGTGTACAGCAATGTTTAAAGAGTGAACAGCC 2517
QY 2565 --TGCAACCGGTGGCCACTCTGTCTTCTTACTGTGAGAGTTCCTTTAAAAAACAACACTGG 2622
Db 2518 CTGCAAGCCCGTGGCCACTCTGTCTTCTTACTGTGAGAGTTCCTTTAAAAAACAACACTGG 2577
QY 2623 CAGCTGCTCTTGGACATGTTTTTAAGAAACAACACTGTATCT 2664
Db 2578 CAGCTGCTCTTGGACATGTTTTTAAGAAACAACACTGTATCT 2619

```

```

RESULT 3
AK086606
LOCUS AK086606 1342 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:D930040I07 product:hypothetical protein, full insert sequence.
ACCESSION AK086606
VERSION AK086606.1 GI:26352128
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED 11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED 11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 585-590 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1342)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

```

```

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, K., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
Location/Qualifiers
source
1. 1342
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:D930040I07"
/db_xref="taxon:10090"
/clone="D930040I07"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="15 days embryo"
214. _>1341
/notes="unnamed protein product; hypothetical protein (evidence: ProCrest)
putative"
/codon_start=1
/protein_id="BAC39701.1"
/db_xref="GI:26352129"
/tranlation="METGTHARKRPGRLGWSFRLPFLRRSHACSEFPFPPSSRRPN
GNSALPERTRYTKLLSOLLALPSLFOKLLWSOLFGLIPTRLDFAASYALRA
LGRRESAAPTVOKSSKLSDSEDLVSSLLDWREGLQWQSSDLELKLKAQERA
LDSAAPTFLLEQQLWGVLLPSSLQAGLVSHRELSGSLVSGNFKVSYLLN
PSYLDYLPQGLRCQSSAGGQVGFRTLTPTSCYLSGDCGHPQPLRAENSAWRRRC
PPLSTGLPBIHHLRMKRLFLQAKGQELPTPDQNGYHSLSEHNLRLMRDPQHCTD
NPAQAVSPAADREPTEKKPELVIQEVSQSPGSSSLFCBLPV"
Query Match 23.2%; Score 1271; DB 3; Length 1342;
Best Local Similarity 98.5%; Pred. No. 1,1e-307;
Matches 1325; Conservative 0; Mismatches 15; Indels 5; Gaps 4;
Qy 250 GCACACCGCTTGGCGGGCCCTCTGGCGCTGCGGCTGCGACACTCGTTGCGGAAGCCGCGC 309
Db 2 GCCACCGCTTGGCGGGCCCTCTGGCGCTGCGGCTGCGACACTCGTTGCGGAAGCCGCGC 61
Qy 310 TCTCTGGCCCTCTCTGTCGCGCGCGGAATCGGACTGCACTACCACTCCGCTGCTGGC 369
Db 62 TCTCTGGCCCTCTCTGTCGCGCGCGGGAATCGGACTGCACTACCACTCCGCTGCTGGC 121
Qy 370 AAGCGGAGACTGTGTAGACTCGGATCCAGCTGCGCTGCGCTGACCGCTGAGCTGTGCT 429
Db 122 AAGCGGAGACTGTGTAGACTCGGATCCAGCTGCGCTGCGCTGACCGCTGAGCTGTGCT 181
Qy 430 CCTCTCTGTGAGAACCGCCCAAGGAAGAGATGGAGACAGGAACCCACAGGCCCGGA 489
Db 182 CCTCTCTGTGAGAACCGCCCAAGGAAGAGATGGAGACAGGAACCCACAGGCCCGGA 241

```

```
Qy 490 AGCGCCTGGCCCTCGGCTGGGCTCCTGGTCCGGCTGCCCTTCTTCCGGGAGTCGACG 549
Db 242 AGCGCCTGGCCCTCGGCTGGGCTCCTGGTCCGGCTGCCCTTCTTCCGGGAGTCGACG 301
Qy 550 CTTGCTCTTCGGAGTTCCTCGGCTGGGCTCCTCTCGACAAAATCCCGGAACTCCGCTCTGC 609
Db 302 CTTGCTCTTCGGAGTTCCTCGGCTGGGCTCCTCTCGACAAAATCCCGGAACTCCGCTCTGC 361
Qy 610 CCGAGCTCGGACGAGTTCGAGCAAAATTCCTCTCGAGCTCTTCCGCTGCCCTCCCTA 669
Db 362 CCGAGCTCGGACGAGTTCGAGCAAAATTCCTCTCGAGCTCTTCCGCTGCCCTCCCTA 421
Qy 670 GCCTATTCCAGAAGCTGCTTTCGAGCCAGCTTTCGGGGGCTGATTCCTTACCAGAT 729
Db 422 GCCTATTCCAGAAGCTGCTTTCGAGCCAGCTTTCGGGGGCTGATTCCTTACCAGAT 481
Qy 730 GGCTAGATTTTCGCGCAAGTTACAGCGCCCTGAGAGCTTCGAGAGAGCGGAGAACTG 789
Db 482 GGCTAGATTTTCGCGCAAGTTACAGCGCCCTGAGAGCTTCGAGAGAGCGGAGAACTG 541
Qy 790 ACCTGCCACGGTCAGAGTCTCTGAGTTACACTCGGCTGGACTCTTCGCGAAGACTC 849
Db 542 CCCTGCCACGGTCAGAGTCTCTGAGTT-CACTGGGCTGGACTCTTCGGAAGACTT- 599
Qy 850 CGCTCGTCAGTACTCTTGCATTGGCTAGAGGAGGACTCCAGTGGCAGTGCTCGTCTCA 909
Db 600 -GGTGTGCTGCTCTTG-ATTGGCTAGAGGAGGACTCCAGTGGCAGTGCTCGTCTCA 657
Qy 910 GA-CTGGAAGTTAACTCAAGGCCAGAAAGAGCTTTAGACTCTGAGCGCCCACTTTC 968
Db 658 GACCTGGAGTTAAACTCAAGGCCAGAAAGAGCTTTAGACTCTGAGCGCCCACTTTC 717
Qy 969 CTCTGAGCAGAGCTGCTGGGAGTGGAGTTGCTGCCAGTACTTCAGCTGGTCTA 1028
Db 718 CTCTGAGCAGAGCTGCTGGGAGTGGAGTTGCTGCCAGTACTTCAGCTGGTCTA 777
Qy 1029 GTCTCCCAACGAGAACTTACTCTTCTCTCTGGGCTCTGAGCGTTCAGAGCTTAGGT 1088
Db 778 GTCTCCCAACGAGAACTTACTCTTCTCTGGGCTCTGAGCGTTCAGAGCTTAGGT 837
Qy 1089 AATTTCAAGGTAGTTTCTATCTCTGAACCTTCTTCACTGAGTACTACTTCCCAAGTTA 1148
Db 838 AATTTCAAGGTAGTTTCTATCTCTGAACCTTCTTCACTGAGTACTACTTCCCAAGTTA 897
Qy 1149 GGGCTGGCTGTCAGAGCGCTGGAGGTGGCCAGTTTGGGTTTTCGAGAACTAACC 1208
Db 898 GGGCTGGCTGTCAGAGCGCTGGAGGTGGCCAGTTTGGGTTTTCGAGAACTAACC 957
Qy 1209 CCAGAGAGCTGCTATCTTCTGAAGATGGTTGTCAACCTCAGCGCTTGGGGCAGAGATG 1268
Db 958 CCAGAGAGCTGCTATCTTCTGAAGATGGTTGTCAACCTCAGCGCTTGGGGCAGAGATG 1017
Qy 1269 TCGGCAACCGCTCGAGAGGTGTCGCGCTCTCTCTACAGAAAGCGCTTCCGGAATCCAC 1328
Db 1018 TCGGCAACCGCTCGAGAGGTGTCGCGCTCTCTCTACAGAAAGCGCTTCCGGAATCCAC 1077
Qy 1329 CACCTTGTATGAACCGCTAGATTCCTCCAGGCTTAAAGGGCAGAGTTTACCAC 1388
Db 1078 CACCTTGTATGAACCGCTAGATTCCTCCAGGCTTAAAGGGCAGAGTTTACCAC 1137
Qy 1389 CCTGCAAGATTAATGGCTATCATAGCTGGAGGAGAACTAACCCTTCTCCGATGGAC 1448
Db 1138 CCTGCAAGATTAATGGCTATCATAGCTGGAGGAGAACTAACCCTTCTCCGATGGAC 1197
Qy 1449 CCAACAATTCACAGATAACCCAGCAGCGGTGTCCTCTGTCAGACAGCGCCGAG 1508
Db 1198 CCAACAATTCACAGATAACCCAGCAGCGGTGTCCTCTGTCAGACAGCGCCGAG 1257
Qy 1509 CCCACTGAGAAAAACCGAATTTGGTATTCAGAGTTTTCACAGAGCCCGGAGGAGC 1568
Db 1258 CCCACTGAGAAAAACCGAATTTGGTATTCAGAGTTTTCACAGAGCCCGGAGGAGC 1317
```

```
Qy 1569 AGTCTGTTTGTGAATTACCCGTGG 1593
Db 1318 AGTCTGTTTGTGAATTACCCGTGG 1342

RESULT 4
LOCUS BQ922494
DEFINITION AGENCOURT 8921919 NCI CGAP Co24 Mus musculus cDNA clone
IMAGE:6395811 5', mRNA sequence.
ACCESSION BQ922494
VERSION BQ922494.1 GI:22337525
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13891 row: b column: 04
High quality sequence stop: 594.

FEATURES
source
1..871
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6395811"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

ORIGIN
Query Match 15.1%; Score 828.2; DB 5; Length 871;
Best Local Similarity 97.9%; Pred. No. 2.2e-196;
Matches 839; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 4463 TTCACTTAGCGTCACTGTTGTCATGTTGAGTAACTGTTACCACTTTTAGTGGT 4522
Db 1 TTGTTGTTGTTGTCATGTTGTCATGTTGAGTAACTGTTACCACTTTTAGTGGT 60
Qy 4523 TACATGCACAAAAGAGAGGACCTCTGAGTCGGGTGGGATGAGCTTTCAGACCTGGCA 4582
Db 61 TACATGCACAAAAGAGAGGACCTCTGAGTCGGGTGGGATGAGCTTTCAGACCTGGCA 120
Qy 4583 GGGTAAACTACCTCAGTTTATAATCTCCCTGGTTATTTCCTGTTGATGTGATTAAGGTC 4642
Db 121 GGGTAAACTACCTCAGTTTATAATCTCCCTGGTTATTTCCTGTTGATGTGATTAAGGTC 180
Qy 4643 TGCTCAGTGGTGTATGATGTTTCATCCACACACAAAGGTTAGTAAGTGCACGACGAGAAA 4702
Db 181 TGCTCAGTGGTGTATGATGTTTCATCCACACACAAAGGTTAGTAAGTGCACGACGAGAAA 240
Qy 4703 CGTTGGTCTTATTTTGGAGAACCCCAATTCCTGTGTTATTTATGCACTCCCTTTAGTGA 4762
Db 241 CGTTGGTCTTATTTTGGAGAACCCCAATTCCTGTGTTATTTATGCACTCCCTTTAGTGA 300
Qy 4763 ACTCCAGAGTGCAATTAAGAGTCTGGTTTAGTCCCGTGGGAATGGGCTAGTTTAGAGCT 4822
```


Db 301 ACTCCAGAGTGCATTAAGAGCTCTGGTTTAGTGCCGTGGGAATGGGCTAGTTTAAAGCT 360
Qy 4823 ATGTTTGAAGACGAGCAAGTTGACTTTAGGAAGAAAAGCTGTGACAGTGTGTAGACATT 4882
Db 361 ATGTTTGAAGACGAGCAAGTTGACTTTAGGAAGAAAAGCTGTGACAGTGTGTAGACATT 420
Qy 4883 TCTTTTAAACCGGACTGCGAGCTTAAACAACCTGTATTTTTCAGATGATTAGGTTTGTTC 4942
Db 421 TCTTTTAAACCGGACTGCGAGCTTAAACAACCTGTATTTTTCAGATGATTAGGTTTGTTC 480
Qy 4943 TGAGACCCAGCACTGTATATTTTAAATTTTTCAGATGATTAGGTTTTCAGATGATTAGGTTTTC 5002
Db 481 TGAGACCCAGCACTGTATATTTTAAATTTTTCAGATGATTAGGTTTTCAGATGATTAGGTTTTC 540
Qy 5003 GTAAATGACTCATGCTCAGACATGCTCTGATGCTGCAAGAAAGCAAGATCTTTGAC 5062
Db 541 GTAAATGACTCATGCTCAGACATGCTCTGATGCTGCAAGAAAGCAAGATCTTTGAC 600
Qy 5063 TGAAGGAGAAAACCTGTCTATTCAGGCTGCTGATGCTGCAAGAAAGCAAGATCTTTGAC 5122
Db 601 TGAAGGAGAAAACCTGTCTATTCAGGCTGCTGATGCTGCAAGAAAGCAAGATCTTTGAC 660
Qy 5123 CAGGACGAGCAGCAGGATGCTGCTGATGCTGCAAGAAAGCAAGATCTTTGAC 5182
Db 661 CAGGACGAGCAGCAGGATGCTGCTGATGCTGCAAGAAAGCAAGATCTTTGAC 720
Qy 5183 CTTTAGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGGCCACACTTCTGAGGGCTG 5242
Db 721 CTTTAGATCTGTGTAGCATGTGAGGCTCTGTACAGTGGGGCCACACTTCTGAGGGCTG 780
Qy 5243 AAATGTGCAACCCCTTATCTAAGCTTGAATCAAAACCGTCAAAATTTTATTAAT 5302
Db 781 AAATGTGCAACCCCTTATCTAAGCTTGAATCAAAACCGTCAAAATTTTATTAAT 840
Qy 5303 TTAAGAAAGAGTGGG 5319
Db 841 TTTAGAAAAGAGTGG 857

RESULT 5
BQ926670
LOCUS
DEFINITION BQ926670 966 bp mRNA linear EST 20-AUG-2002
IMAGE:6398165 5', mRNA sequence.
ACCESSION BQ926670
VERSION BQ926670.1 GI:22341701
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
cDNA Library Preparation: The Cepko Laboratory
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (ILNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
<http://image.llnl.gov>
Plate: LLNL13897 Row: d Column: 06
High quality sequence stop: 613.
Location/Qualifiers
1..966
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6398165"

FEATURES
source

/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Query Match 14.8%; Score 810; DB 5; Length 966;
Best Local Similarity 96.4%; Pred. No. 8.3e-192;
Matches 873; Conservative 0; Mismatches 25; Indels 8; Gaps 4;
Qy 3935 GAAAGCGGGTGGTGTCTTTTCAAGGACACACCATTTATGCCAGCTGGTCTGCCAA 3994
Db 1 GAAAGCGGGTGGTGTCTTTTCAAGGACACACCATTTATG-CGAGCTGGTCTGCCAA 59
Qy 3995 ACTTAAACATATTTGGTTTCTGTGCAAACTGTCATTTGGAGGTTTCTGTGTGTGTAGT 4054
Db 60 ACTTAAACATATTTGGTTTCTGTGCAAACTGTCATTTGGAGGTTTCTGTGTGTGTAGT 119
Qy 4055 TTCAAGTTGAATGTGGCTCTCTGTAGCTTTCAAGAAATGGGTAGAAATCATAAAGCACCTTT 4114
Db 120 TTCAAGTTGAATGTGGCTCTCTGTAGCTTTCAAGAAATGGGTAGAAATCATAAAGCACCTTT 179
Qy 4115 AAGTAATCATGTGCATGTAGACATTTTAACTAGGGGTTATTTGGGAGACATG 4174
Db 180 AAGTAATCATGTGCATGTAGACA--TTTTTTTAACTAGGGGTTATTTGGGAGACATG 237
Qy 4175 TCAGATATTCACATTTGTTTATGTGCTTTAAACACAGTCTTACTTACACCTGCCTCAG 4234
Db 238 TCAGATATTCACATTTGTTTATGTGCTTTAAACACAGTCTTACTTACACCTGCCTCAG 297
Qy 4235 CACTGGCACCTTTTCAAAACCTGTCTTGGAGACACCTGTAAACTTGGATGGTGCAGTTCCTGG 4294
Db 298 CACTGGCACCTTTTCAAAACCTGTCTTGGAGACACCTGTAAACTTGGATGGTGCAGTTCCTGG 357
Qy 4295 TTTTCATGTGTAAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTTCACTT 4354
Db 358 TTTTCATGTGTAAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTCTTCACTT 417
Qy 4355 ATGTAGAGAGACAGAGCAGAGGACAGATAGAGCAGCAGAGGCTAGTGTCTCC 4414
Db 418 ATGTAGAGAGAGACAGAGCAGAGGACAGATAGAGCAGCAGAGGCTAGTGTCTCC 477
Qy 4415 CCCAAGCGTGTCTTAGATAGTGTGTGATCAGGTGTCTTGGTTTGTTCAGTTAGCGT 4474
Db 478 CCCAAGCGTGTCTTAGATAGTGTGTGATCAGGTGTCTTGGTTTGTTCAGTTAGCGT 537
Qy 4475 CACTGTGTGATGTTGAGGTTAATGTGTACAGTTCTTTAGTGTGTATCATGCACAA 4534
Db 538 CACTGTGTGATGTTGAGGTTAATGTGTACAGTTCTTTAGTGTGTATCATGCACAA 597
Qy 4535 AGAGAGACCTCTGAGTCGGGTGTGGATGAGCTTTCCAGACCTGGCAGGTTAACTACC 4594
Db 598 AGAGAGACCTCTGAGTCGGGTGTGGATGAGCTTTCCAGACCTGGCAGGTTAACTACC 657
Qy 4595 TCAGTTTATAATCTCCCTGGTTATTTCCGTTTGAATGTATTAAGTCTGCTCAGTGGT 4654
Db 658 TCAGTTTATAATCTCCCTGGTTATTTCCGTTTGAATGTATTAAGTCTGCTCAGTGGT 717
Qy 4655 GATGATGTTATCACCACACAGAGTTAGTAAAGTGCAGACAGAAACGTTGTCTTAT 4714
Db 718 GATGATGTTATCACCACACAGAGTTAGTAAAGTGCAGACAGAAACGTTGTCTTAT 777
Qy 4715 TTTTGAAGACCCCATTTCTGTGTATTTTATGCA-CCTGCTTTTGTAGTAACTCCAGAGTG 4773
Db 778 TTTTGAAGACCCCATTTCTGTGTATTTATGACCCCTGCTTTTGTAGTAACTCCAGAGTG 837
Qy 4774 CATTTAAAGAGTCTGG-----TTTGTGCTGGGAAATGGGCTAGTTTAAAGTATGTTG 4829
Db 838 CATTTAAAGAGTCTGGGTTTAAATGGCTGGGAAATGGGCTACTTTACAAACATTTT 897
Qy 4830 GAAAGC 4835

Db	898	GGAAAC	903		/dev_stage="adult" /clone_lib="Mus musculus pancreas C57BL/6J adult"
RESULT 6	AV055059				
LOCUS	AV055059	997 bp	mRNA	linear	EST 24-OCT-2001
DEFINITION	AV055059	Mus musculus pancreas C57BL/6J adult	Mus musculus	cdna	
ACCESSION	AV055059				
VERSION	AV055059.2	GI:16380666			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Mus musculus				
REFERENCE					
AUTHORS	1 (bases 1 to 997)				
	Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, P., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.				
TITLE	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)				
JOURNAL	Unpublished (2001)				
COMMENT	On Jun 23, 1999 this sequence version replaced gi:5154806. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				
FEATURES	Location/Qualifiers				
source	1..997				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="1810033K10"				
	/sex="male"				
	/tissue_type="pancreas"				
ORIGIN					
Query Match	14.8%	Score 808.6	DB 1	Length 997	
Best Local Similarity	96.4%	Pred. No. 1.9e-191			
Matches 834	Conservative 0	Mismatches 30	Indels 1	Gaps 1	
Qy	4584	GGTAAACTACCTCAGTTTAAATCTCCCTGGTTATTTCCGTTTGATGTGATCTAAGGTCT	4643		
Db	134	GGTAAATCTACCTCAGTTTAAATCTCCCTGGTCAATCCGATTGAATGTATCTAAGTACT	193		
Qy	4644	GCCTCAGTGTGATGATGATTCATCCACACACAGAGGTAGTAAGAGTGCACGACGAGAAC	4703		
Db	194	CCCTC-GTGTGTGATGATGATCTCAACACACAGAGGTAGTAAGAGTGCACGACGAGAAC	252		
Qy	4704	GTGGGTCTTATTTTGGAGAACCCCAATTTCTGTGTATTTTATTCACCTGCCTTTAGTGAA	4763		
Db	253	GTGTGTTCTTATTTGTAGAACCCCAATTTCTGTGTATTTTATTCACCTGCCTTTAGTGAA	312		
Qy	4764	CTCCAGAGTGCATTTAAAGAGTCTGGTTAGTGCCTGGGAATCGGCTAGTTTAAAGACTA	4823		
Db	313	CTCCAGAGTGCATTTAAAGAGTCTGGTTAGTGCCTGGGAATCGGCTAGTTTAAAGACTA	372		
Qy	4824	TGTTTGGAAAGCAGGCAAGTTGACTTTAGGAAGAAAGCTGTGCACAGTGTGTAGACATTT	4883		
Db	373	TGTTTGGAAAGCAGGCAAGTTGACTTTAGGAAGAAAGCTGTGTAGACATTT	432		
Qy	4884	CTTTTAAACCGGAGCTGCAGCTTAAACACACTTCAATTTTCAAGTATAGGTTTTGTTTCT	4943		
Db	433	CTTTTAAACCGGAGCTGCAGCTTAAACACACTTCAATTTTCAAGTATAGGTTTTGTTTCT	492		
Qy	4944	GAGACCCAGCACCTGTATATTTTAAATAATTTTCCAGATTACACCTTCAATCAATGAG	5003		
Db	493	GAGACCCAGCACCTGTATATTTTAAATAATTTTCCAGATTACACCTTCAATCAATGAG	552		
Qy	5004	TAAATGACTCATGCTGCAGACATGCTGATGGTGGCAAGAACAGAGATCTTTGACT	5063		
Db	553	TAAATGACTCATGCTGCAGACATGCTGATGGTGGCAAGAACAGAGATCTTTGACT	612		
Qy	5064	GAAGGAGAAAACGTGTCATTTCCAGCTTCAATTTTCCAGATTACACCTTCAAGGCGAGGC	5123		
Db	613	GAAGGAGAAAACGTGTCATTTCCAGCTTCAATTTTCCAGCTTCAATTTCAAGGCGAGGC	672		
Qy	5124	AGGCGAGCAGGCGGATGCTGTTCTTAGTTGAATACACATTTCAAGTCTTGCAGTGTGC	5183		
Db	673	AGGCGAGCAGGCGGATGCTGTTCTTAGTTGAATACACATTTCAAGTCTTGCAGTGTGC	732		
Qy	5184	TTTAGATCTGTGATGATGAGGCTCTGTACAGTGGGGGCCACACTTCTGAGGGCTGA	5243		
Db	733	TTTAGATCTGTGATGATGAGGCTCTGTACAGTGGGGGCCACACTTCTGAGGGCTGA	792		
Qy	5244	AATGTGGCAACCTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTATTAATT	5303		
Db	793	AATGTGGCAACCTTTATCTAACTTGAATCAAAACCGTCAAAATTTTATTTTATTAATT	852		
Qy	5304	TAAGAAGAGTGTGGGAATGACATTTTTCAGTTGGCTTTTTCAGCTCAGTCAATTTACG	5363		
Db	853	TAAGAAGAGTGTGGGAATGACATTTTTCAGTTGGCTTTTTCAGCTCAGTCAATTTACG	912		
Qy	5364	TGTAACGTGAGATTTGATAGTCAAGTATATTTGTATATATTAATTAATCTAATCTGTA	5423		
Db	913	TGTAACGTGAGATTTGATAGTCAAGTATATTTGTATATATTAATTAATCTAATCTGTA	972		
Qy	5424	AATTGTAATAAATATATTTTGCRAAT	5448		
Db	973	AATTGTAATAAATATATTTTGCRAAT	997		
RESULT 7	CF583619	890 bp	mRNA	linear	EST 24-SEP-2003
LOCUS	CF583619/c				
DEFINITION	AGENCOURT_0845828_updated NIH_MGC_137 Mus musculus				
	cdna clone				

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11) 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, J., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 e mouse tissues.

FEATURES

Location/Qualifiers
 1. .1011
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1810033K10"
 /sex="male"
 /tissue_type="pancreas"
 /dev_stage="10 day old"
 /clone_lib="RIKEN full-length enriched, 10 day old male pancreas"

ORIGIN

Query Match 14.6%; Score 796; DB 2; Length 1011;
 Best Local Similarity 90.9%; Pred. No. 2.8e-188;
 Matches 924; Conservative 0; Mismatches 83; Indels 10; Gaps 7;
 3724 TTTCTTTTTTAAAGACATGGTCTTATATAGTCTAGGCTGGCCTTAACTCATTA 3783
 Db 1 TTTCTTTTTTAAAGACATGGTCTTATATAGTCTAGGCTGGCCTTAACTCATTA 60
 3784 GCCAAGATGGCTTTGAAGATCTCTGCTGCTCTTCTTGGCTGACATGGCATGTC 3843
 Db 61 GCCAAGATGGCTTTGAAGATCTCTGCTGCTCTTCTTGGCTGACATGGCATGTC 120
 3844 CATACATCCAGATTTTCTCTGTATCTAGGCTCTTTTAAAGATATTTCCAAAGACTGCCCA 3903
 Db 121 CATACATCCAGATTTTCTCTGTATCTAGGCTCTTTTAAAGATATTTCCAAAGACTGCCCA 180
 3904 CAAATCCACAGTCAGTACTTTCTTTGACTGGGAAAGCGGGTGGTGTCTTTTCAAAGG 3963
 Db 181 CAAATCCACAGTCAGTACTTTCTTTGACTGGGAAAGCGGGTGGTGTCTTTTCAAAGG 240
 3964 CACACACATATGCCAGCTGGTCTGCCAACTTAACATATTTGGTTCTGTGCAACT 4023
 Db 241 CACACACATATATGCCAGCTGGTCTGCCAACTTAACATATTTGGTTCTGTGCAACT 300
 4024 GTCCATTTGGAAGGTTCTGTGTCTGTGTAGTTTTCAGTTGATGGTCTCTGTAGCTTT 4083
 Db 301 GTCCATTTGGAAGGTTCTGTGTGTGTGTAGTTTTCAGTTGATGGTCTCTGTAGCTTT 360
 4084 CAAGAATGGGTAGAAATCATAAAGCACTCTTTAAGTAATCAATTCATTTGACATTTTTT 4143
 Db 361 CAAGAATGGGTAGAAATCATAAAGCACTCTTTAAGTAATCAATTCATTTGACATTTTTT 419
 4144 TTTTAACTAGGGGGTATTTGGGAGACATGTACATATTCACATTTTGTATGTGTCT 4203
 Db 420 TTTTAACTAGGGGGTATTTGGGAGACATGTACATATTCACATTTTGTATGTGTCT 479
 4204 TAAACACAGTCTTACTTACACCTGCTCAGCACTGGCCTTTTCAAACCTGTCTTGGAG 4263
 Db 480 TAAACACAGTCTTACTTACACCTGCTCAGCACTGGCCTTTTCAAACCTGTCTTGGAG 539
 4264 ACACGTAAACTTGGATGGTGTGAGTCTCTGGGTTTTTCATGTGTAAATGTCAACTACAAAG 4323

Db 540 ACACGTAAACTTGGATGGTGCAGTTCTGGGTTTTTCATGTGTAAATGTAACTACAAAG 599
 Qy 4324 GTCAAATATCCAGGTTTGTGTGTCTTACCTTATGTAGAGAGAAAGCAAGCAAGGGCA 4383
 Db 600 GTCATATATCCAGGTTTGTGTGTCTTACCTTATGTAGAGAGAAAGCAAGGGCA 659
 Qy 4384 GATAGAGCAGCCCAAGAGTGTCTAGT--GTCCCCCCCCAAAGCGGTCTTCTAGATAGTGTGT 4441
 Db 660 GATAGAGCAGCCAGATAGTGTCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 719
 Qy 4442 GGATCAGGT 4501
 Db 720 GGATCAGGT 779
 Qy 4502 TGTTCACAGTCTTTTGT 4561
 Db 780 TGTGTACAGTCNTTTAGGGTTACGACACAAAGAGA--GACCTCTGAGTCGGGTGTGGG 836
 Qy 4562 ATGAGCTTTTCCAGACCTGGCAGGGTAACTTACCTCAGTTTATATCTCCTCTGTGTATTTTC 4621
 Db 837 GATGAGCTTCCAGACACACGCGGAGGTAGACTATCTCAGTTAAT-ATCTCCCTGTATATT 895
 Qy 4622 CGTTTGTATGTATTAAGTCTGCTCCTCAGTGGTGTATGATGTTTCATCCACACAAAGTTA 4681
 Db 896 CGCTTGAAGCGATCTAATG-CTGCTCAGTCAGTCGTGAAGACGTTTCAT-CACACACAAAGCTA 953
 Qy 4682 GTAAGAGTGCACACACAGAAACGTTGTGTCTTATTTTGTAGAACCCCATTTCTGTGTGT 4738
 Db 954 GNAATAGTCACGA-CAGAGAAGTGGCTATTTTATAGNACCCCTCTCTGTGTGT 1009

RESULT 9
 CF583620
 LOCUS
 DEFINITION
 CF583620
 VERSION
 CF583620.1 GI:35196882
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC <http://mgi.nci.nih.gov/>
 1 (bases 1 to 913)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs@mail.nih.gov
 Tissue procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)
 cDNA Library Preparation: Catherine Lee, Endocrine Pancreas Consortium
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Cloning by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: IRBD20 row: h column: 10
 High quality sequence stop: 672.

FEATURES

Location/Qualifiers
 1. .913
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6431877"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_137"
 /notes="Organ: pancreas; Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Library consists of a pool of clones"


```

Qy 3616 TCTCCCTC-TTTTCTTTTTCCTCCAGAGCTGAGGACAAACCAGGCTTTGACCTTGC 3674
Db 423 TCTCCCTC-TTTTCTTTTTCCTCCAGAGCTGAGGACAAACCAGGCTTTGACCTTGC 482
Qy 3675 TAGGCAAGCGCTTACCACTGAGCTAAATCCCAAAACCAACCCGCCCTTTCTTTT 3734
Db 483 TAGGCAAGCGCTTACCACTGAGCTAAATCCCAAAACCAACCCGCCCTTTCTTTT 541
Qy 3735 TAAAGACATGCTTATATAGTCTAGGCTGGCCCTTAAACTCATTTAGCCAAAGGATGG 3794
Db 542 TAAAGACATGCTTATATAGTCTAGGCTGGCCCTTAAACTCATTTAGCCAAAGGATGG 601
Qy 3795 CTTTGAAGATCTCTGCTCTTCTTTGCTTGGCATGCGCATGCGCATCATCCAG 3854
Db 602 CTTTGAAGATCTCTGCTCTTCTTTGCTTGGCATGCGCATGCGCATCATCCAG 661
Qy 3855 ATTTTCTCTGTATCTAGTCTTTTAAAGATATTTCCAAAGGACTGCCCAAAATCCACAG 3914
Db 662 ATTTTCTCTGTATCTAGTCTTTTAAAGATATTTCCAAAGGACTGCCCAAAATCCACAG 721
Qy 3915 TGCAGTACTTTCTTGAAGCTGGG-AAAGCGGGGTTGGTGTCTTTTCAAAGGCACACCAT 3973
Db 722 TGCAGTACTTTCTTGAAGCTGGGAAAGCGGGTGGTGTCTTTTCAAAGGCACACCAT 781
Qy 3974 TATGCCAGCTGCTGCTGCTTGTGCTTGGCATGCGCATGCGCATCATCCAG 4032
Db 782 TATGCCAGCTGCTGCTGCTTGTGCTTGGCATGCGCATGCGCATCATCCAG 841
Qy 4033 GAAGTCTCTGTGT-GTTGTAGTTTTCAGTTGAATG-TGGCTCTCTGAGCTTTCAG--A 4088
Db 842 GAAGTCTCTGTGTGTTGTTGGAATTTTCAGTTGAATGTTGGCTCTCTGAGCTTTTAAAGAA 901
Qy 4089 ATGGGTAGAAATCATAAGCA 4109
Db 902 AGGGTAAAGTCTTAAAGCA 922

RESULT 11
HSM805059 3941 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFP6661186 (from clone DKFP6661186).
DEFINITION AL833746
VERSION AL833746.1 GI:21734396
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German cDNA Consortium
Direct Submission
Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFP6661186) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFP6661186
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.
Location/Qualifiers
1. .3941
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFP6661186"
/db_xref="taxon:9606"
/clone="DKFP6661186"

FEATURES
source

```

```

/tissue type="stomach"
/clone_lib="666 (synonym: hsto2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev stage="adult"
/notes="hypothetical protein, N-terminus truncated"
1. .3941
/gene="DKFP6661186"
<2. .1225
/gene="DKFP6661186"
/codon start=1
/product="hypothetical protein"
/protein_id="CAH56240.1"
/db_xref="GI:52545825"
/translation="ASKGDLPTPDQDNGYHSLBEEHSLRMDPKHCRDNPOTQVPAA
GDTGNTQSTEEKIELLTTEVLALEESPSGCSFSEIPEKFEGRGEGISVVDYSY
LEGDLPIASRACSNKLDIYILGASDLETSDEPEDWDEAEADDFDSDSLSDS
DLBQDPGLHLWNPSFCVPYNPONTATITOTAAIRVPEPSPSEKDLGSKSLSDS
QSGLPETPEHSGEEDWESSADEASLKNWFCNSDDPYNPLNPKAPPTQSGENE
KGRDSTKTSIESIVALISECHLLSKVQLGSOSECPDPSVQRDLVLSGGHHTHVKKRK
VTLEBVTYYISGDEDRKGPWBEFARDGCRFQRIQETEDAIGYCLTTFEHRERMFNR
LQTCFKGLNLVKQC"
ORIGIN
Query Match 14.4%; Score 786.4; DB 3; Length 3941;
Best Local Similarity 75.2%; Pred. No. 9.6e-186;
Matches 1108; Conservative 0; Mismatches 311; Indels 55; Gaps 8;
Qy 1361 GGCTAAACAAAGGCAAGAGTTACCCACCCCTGACCAAGATAATGGCTATCATAGCTTGA 1420
Db 1 GGCTAGCAGGGGCAAGATTACCCACCCCTGACCAAGATAATGGCTATCATAGCTTGA 60
Qy 1421 GGAGGAACATAAATCTTCCGGATGGACCCACCAACATTCACAGATAAACCCAGCAGGC 1480
Db 61 GGAGGAACACAGCTTCTCCGGATGGATCCAAACACCTGCAGAGATAAACCCACACAGTT 120
Qy 1481 GGTGTCCCTGCTGCAGACAGGCC-----GGAGCCCACTGAGAGAAAAACAGCA 1528
Db 121 TGTTCCTGCTGCTGGAGACATTCCTCGAAACACCCAGGAATCCCACTGAAGAAAAATAGA 180
Qy 1529 ATTGGTGTATCAAGAAATTTTAC-----AGAGCCCCCAGGGAAGAGCTCT 1573
Db 181 ATTATTAACTACAGAGTTTCCACTTGTGTTGGAAGAGAGAGAGCCCTCTGAGGGCTGTCC 240
Qy 1574 GTTTTGTGAATTAACCCGTGGAAAAAGAAAGTGTGAAGAGGACCAACATAATCAACTGACCT 1633
Db 241 ATCTAGTGAATACCTATATGAAAAGAGAGCCCTGGAGAGGGCCGAATAAGTGTAGTTGATTA 300
Qy 1634 CTGATAGAGGAGAGAGCCCTTCTGTTTCTACAGACCAAGTTTGTAGCAACAACTGAT 1693
Db 301 CTCATACCTAGAAAGGTGACCTTCCCATTTCTGCCAGACCAAGCTTGTAGTAACAACTGAT 360
Qy 1694 AGATTATATTTTGGGAGGGCCGCCAGTGTGCTTTGGAAGCCAGCTCTGATTCTGAAAGTGA 1753
Db 361 AGATTATATTTTGGGAGGTGCATCCAGTGTACCTGGAACCAAGTCTGTATCCAGAGGTGA 420
Qy 1754 GGATTGGGGGAGAAACCTGAGGACGATGGCTTTTGTAGCGATGGCTCCCTGTCTGAATC 1813
Db 421 GGATTGGGATGAGAAAGCTGAGGATGATGATGTTTTGTAGTGTAGTCACTGTGACACTC 480
Qy 1814 AGACGTGGAACAGAGCTCGGAAGCCCTTCACTTTTGGAACTCTTTTCCAGAGTGTAGATCC 1873
Db 481 AGACCTTGAACAAAGACCCCTGAAGGGCTTCACTTTTGGAACTCTTTCTGCAAGTGTAGATCC 540
Qy 1874 TTACAAACCCCAAACTTTTACAGCCACGATTACAGAGCGCTGCCAGAAATGSCCCCAAGAGA 1933
Db 541 TTATAATCCCAAGACTTTTACAGCAAAATTCAGATCTGTGCCAGAAATGTTCTCCTGAAGA 600
Qy 1934 CCCATCAGATTTCAGGGACATCTCTGGTCTGGCAGCTGTGGTGTAGGGA---GCTGTGAGCA 1990
Db 601 GCCTTCTGATTACAGAGAGGATTTGTCTGGCAGTCTGTATCTAGAGAAATTCCTCCAGATC 660
Qy 1991 GGACACCTTCCGGAGACCCCGACCATAGTTTCCGGGGAGGAAGATGACTGTGGGAACCGAG 2050

```

```

Db      661  TGAAGCCTCTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGAATCTAG 720
Qy      2051  TGCAGATCAAGCAGAGATCTTAATCTGGAACCTCTTCTGTCATCTGAGGACCCCTA 2110
Db      721  TGCAGATGAAGCAGAGAGTCTCAAACTGTGGAACCTCTTGTAACTCTGATGACCCCTA 780
Qy      2111  CAACCTTTTAAATTTTAAAGGCTCTTTTCAACCGCTCAGGGAAGAATTTGGAAGGCCGTCA 2170
Db      781  CAACCTTTTAAATTTTAAAGGCTCTTTTCAACATCAGGGGAATGAGAAGGCTGTG 840
Qy      2171  GGACTCAAGGCTCTTCTGAGGTACAGTGGCTTCTCTGCGCATCATCTTACTTTTC 2230
Db      841  TGACTCAAGACCCCATCTGAGTCAATTTGTGGCCATTTCTGAGTGTCAACCTTACTTTTC 900
Qy      2231  TTGTAAGGCCAGCTGTTAGAGAGCCAGAGATAATTTGTCAGGCTGTGGCTGGGTGA 2290
Db      901  TTGTAAGGTGAGCTGTTTGGGAGCCAAAGAAGTGAATGTCCAGACTCGGTACAGCGTGA 960
Qy      2291  GGCTCTTCTCTGGAGAAAGATACCCCATATCAAGAGAAAAAGGTAACTTCTCTGGAAGA 2350
Db      961  CGTCTTCTCTGGAGGAAGACACACATGTCAAAAGAAAAAAGGTAACTTCTCTTGAAGA 1020
Qy      2351  AGTTACTGAGTATTATATAAGTGTGATGAGATGCGAAAGACCATGGGAAGAAATTTGC 2410
Db      1021  AGTTACTGAGTATTATATAAGTGTGATGAGATGCGAAAGACCATGGGAAGAAATTTGC 1080
Qy      2411  AAGGGATGGATGAGGTTTCCAGAACGAAATTCAGAAACAGAAAGTTGCCATTGGCTACTG 2470
Db      1081  AAGGGATGGATGAGGTTTCCAGAACGAAATTCAGAAACAGAAAGTTGCTATTGGATATTG 1140
Qy      2471  CTTGGCTTTTGGACACAGAGAAAAAATGTTTAATAGACT----GAGGATCGAGTCAAAAG 2526
Db      1141  CTTGACATTTGAACACAGAGAAAGAAATGTTTAATAGACTCCAGGGAACATGCTTCAAAG 1200
Qy      2527  ACTTACTGTTGACAGAAATGTTAAGAAGTGA-ACAGCTGCAACCCGTGCCACTCTGT 2585
Db      1201  ACTTAAATTTCTCAAGCAATTTGAGTTGGCAGCCTGTAGTCTTGTAGTATACATAC 1260
Qy      2586  CTCTTACTTGTAGA-GTTTTCCCTTAAAAACAACACTGGCAGCTGTCTTGGACATGTTTT 2644
Db      1261  CTCTTACTGAGAGTGTCTTTTAAAAACAATCTTGGCAGCTGTCTTGGACATTTTTT 1320
Qy      2645  TAAAGAAACAACCTTGTATCTAGAGATGCAGTTTGTATTTATTTTGGGTAATGTCTCAT 2704
Db      1321  TTTTATAGGAAATGTAATCTGATCTAGTTAAATTTTTTTTTTTTGGCAACATATCCACTC 1380
Qy      2705  AGAAACA-----CCACTCCGATAATGAAGAAAT-----CTCTTACTGTAA 2745
Db      1381  AGAAACATTCAGGTTTGAAGCCAGCCCTGTGATAATGAAGGATGAAGTGTGATTCTAA 1440
Qy      2746  TCCTCTCTTTTCTTATTTAGTTGGATGGGTTTT 2779
Db      1441  TCCTCCCTTTTGTATTTAGTTGGATGTGCTTTTT 1474

RESULT 12
LOCUS   CO428800
DEFINITION  UI-M-HX0-crw-c-05-0-UI.r1 NIH_BMAP_HX0 Mus musculus CDNA clone
IMAGE:30686164 5', mRNA sequence.
ACCESSION  CO428800
VERSION    CO428800.1 GI:49675094
KEYWORDS  EST.
SOURCE     Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE  1 (bases 1 to 801)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov

```

Tissue Procurement: Dr. James Lin University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

The following repetitive elements were found in this cDNA
 sequence: 602-707, >BI-FHSINE/Alu (matched complement)
 Seq primer: pYX-5.

FEATURES
 source

Location/Qualifiers
 1..801
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30686164"
 /tissue_type="whole eye"
 /dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_HX0"
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bontal, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag is AATAATTACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Query Match 14.28; Score 774.8; DB 7; Length 801;
 Best Local Similarity 99.4%; Pred. No. 5.6e-193;
 Matches 798; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 2176 CAAAGGCTCTTCTGAGGTACAGTGGCTTCTCTGCGCATCATACCTTCTTTGTA 2235
 Db 1 CAAAGGCTCTTCTGAGGCTACAGTGGCTTCTCTGCGCATCATACCTTCTTTGTA 60

Qy 2236 AGGCCAGCTGTTAGAGAGCCAGAGATAATTTGTCAGGCTGTGGCTGGTGAGGCTC 2295
 Db 61 AGGCCAGCTGTTAGAGAGCCAGAGATAATTTGTCAGGCTGTGGCTGGTGAGGCTC 120

Qy 2296 TTGCTGGAGAAAGATACACCCATATCAAGAGAAAAAGGTAACTTCTTGGAGAGTTA 2355
 Db 121 TTGCTGGAGAAAGATACACCCATATCAAGAGAAAAAGGTAACTTCTTGGAGAGTTA 180

Qy 2356 CTGAGTATTATATAAGTGTGATGAGATCGCAAGGACCATGGGAAGAAATTTGCAAGG 2415
 Db 181 CTGAGTATTATATAAGTGTGATGAGATCGCAAGGACCATGGGAAGAAATTTGCAAGG 240

Qy 2416 ATGATGCGAGGTTCCAGAAACGAATTCAGAAACAGAAAGTTGCCATTGGCTGCTTGG 2475
 Db 241 ATGATGCGAGGTTCCAGAAACGAATTCAGAAACAGAAAGTTGCCATTGGCTGCTTGG 300

Qy 2476 CCTTTGAGACACAGAAAAAATGTTTAAATAGACTGAGGATCGAGTCAAGGACCTTACTGT 2535
 Db 301 CCTTTGAGACACAGAAAAAATGTTTAAATAGACTGAGGATCGAGTCAAGGACCTTACTGT 360

Qy 2536 TGTAACAGCAATGTTAAGAAAGTGAACAGCTGCAACCCGTGCCCACTCTGTCTTACTTG 2595
 Db 361 TGTAACAGCAATGTTAAGAAAGTGAACAGCTGCAACCCGTGCCCACTCTGTCTTACTTG 420

Qy 2596 AGAGTTTCCCTTAAAAACAACACTGCGAGCTGCTTGGACATGTTTTTAAAGAAACAA 2655

Db 421 AGAGTTTCCCTTAAACAAACACATGGCAGCTGCTCTTGGACATGTTTTTAAAGAACAA 480
Qy 2656 CTTGTATCTAGAGATGCAAGTTGATTAATTTTGGGTAATGTCCTCATTTAGAAACACCAA 2715
Db CTTGTATCTAGAGATGCAAGTTGATTAATTTTGGGTAATGTCCTCATTTAGAAACACCAA 540
Qy 2716 CTCGATTAATGAAGATCTCTATCTGTAATCTCTCTTTCCCTATTTAGTTGATGCG 2775
Db CTTCCGATTAATGAAGATCTCTATCTGTAATCTCTCTTTTCTTCTATTTAGTTGATGCG 600
Qy 2776 GTTTTGTCTTTTGGAGGGTCTCACTGCATAATTTCTGTTTGGCTAGGCTATGATGATGT 2835
Db GTTTTGTCTTTTGGAGGGTCTCACTGCATAATTTCTGTTTGGCTAGGCTATGATGATGT 660
Qy 2836 AGACAGGTTGGCCTTGAAGTTGGTAAATTCCTTGCTCTGTTTCTGTCGACCATGCC 2895
Db AGACAGGTTGGCCTTGAAGTTGGTAAATTCCTTGCTCTGTTTCTGTCGACCATGCC 720
Qy 2896 AGCTGGAAGTGTTTTAAATATCTCTGCTTACCTGGGGTGAGAGTTTAAATTTGCACCTT 2955
Db AGCTGGAAGTGTTTTAAATATCTCTGCTTACCTGGGGTGAGAGTTT-ATTTTGCACCTT 779
Qy 2956 TCAAAACGTTTCTTCTGGAGGCA 2978
Db TC-AAACGTTTCTTCTGGAGGCA 801

RESULT 13
LOCUS BI654191 796 bp mRNA linear EST 12-SEP-2001
DEFINITION G03280489P1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5324446 5', mRNA sequence.
ACCESSION BI654191
VERSION BI654191.1 GI:15568427
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 796)
TITLE NIH-MGC <http://mgi.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLM11823 row: a column: 23
High quality sequence stop: 785.
Location/Qualifiers
1. 796
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NWRI"
/db_xref="taxon:10090"
/clone="IMAGE:5324446"
/tissue_type="tumor, gross tissue"
/dev_stages="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam4"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."

FEATURES
source

Query Match 14.0%; Score 767.4; DB 4; Length 796;
Best Local Similarity 99.1%; Pred. No. 4.1e-181;
Matches 782; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1 CGGTCTCTCGCTCTCGCCCTGACGCTTCGGGGTGGCGGCTGCGGCCATTTTGGCTTCGC 60
Db 1 CGGTCTCTCGCTCTCGCCCTGACGCTTCGGGGTGGCGGCTGCGGCCATTTTGGCTTCGC 60
Qy 61 TTCTTTTGGCCCTCGCCCTGCGCACCCAGCCACCCCTTTCCGGCTTTGCGGCTTCGCGCCCTCG 120
Db 61 TTCTTTTGGCCCTCGCCCTGCGCACCCAGCCACCCCTTTCCGGCTTTGCGGCTTCGCGCCCTCG 120
Qy 121 TGGGGCCACCGGAAACCGCCGCTCGTCTCGGTCGCGCGCGCGGAGGAGGCTTCTTCT 180
Db 121 TGGGGCCACCGGAAACCGCCGCTCGTCTCGGTCGCGCGCGCGGAGGAGGCTTCTTCT 180
Qy 181 CTATGTTGGAGCGATCTCACACGGCCTTAGACGCTCTCTTCCCTAGCGGGGATGGACCTA 240
Db 181 CTATGTTGGAGCGATCTCACACGGCCTTAGACGCTCTCTTCCCTAGCGGGGATGGACCTA 240
Qy 241 ACCGCGTTCGCCACCGCTTTCGCGGGCTCTGCGGCCGTCGCGTGCAGACTCGTTGCGGA 300
Db 241 ACCGCGTTCGCCACCGCTTTCGCGGGCTCTGCGGCCGTCGCGTGCAGACTCGTTGCGGA 300
Qy 301 AGCGCGCTCTCTGCGGCTCTTTCGCGCGCGGGAAATCGGACTGCCACTCCG 360
Db 301 AGCGCGCTCTCTGCGGCTCTTTCGCGCGCGGGAAATCGGACTGCCACTCCG 360
Qy 361 TGGCTGGGCAAGCGGAGACTGTGTAGACTTCGGATCCAGCTGCGCTGAGCGCGCTGAG 420
Db 361 TGGCTGGGCAAGCGGAGACTGTGTAGACTTCGGATCCAGCTGCGCTGAGCGCGCTGAG 420
Qy 421 CTCTGTCTCTCTCTGTCTGAGAAAGCGCCCAAGAAAGAGATGGAGACAGGAAACGACA 480
Db 421 CTCTGTCTCTCTCTGTCTGAGAAAGCGCCCAAGAAAGAGATGGAGACAGGAAACGACA 480
Qy 481 GGGCCCGGAAAGCGGCTTTCGAGTTCGCGGCTCTGCGGCTTCGCGTCCCTTCTTCGCG 540
Db 481 GGGCCCGGAAAGCGGCTTTCGAGTTCGCGGCTCTGCGGCTTCGCGTCCCTTCTTCGCG 540
Qy 541 GATCGACGCTGCTCTTCGGAGTTCGCGGCTTCCTCTCGACAAAATCCCGGGAAT 600
Db 541 GATCGACGCTGCTCTTCGGAGTTCGCGGCTTCCTCTCGACAAAATCCCGGGAAT 600
Qy 601 CCGCTCTGCCCGAGCTGCTTCGAGTTCGCGGCTTCCTCTCAGTCTCTTCAGCTCTTCGCC 660
Db 601 CCGCTCTGCCCGAGCTGCTTCGAGTTCGCGGCTTCCTCTCAGTCTCTTCAGCTCTTCGCC 660
Qy 661 TGCTCCCTAGCTTATTCAGAAAGCTGCTGCTTTTGGAGCCAGCTTTCCGGGGGCTGATTC 720
Db 661 TGCTCCCTAGCTTATTCAGAAAGCTGCTGCTTTTGGAGCCAGCTTTTTCGGGGGCTGATTC 720
Qy 721 CTACCAGATGCTAGATTTTGGCGCAAGTTTACAGCGCCCTGAGAGCTTCAGAGGACGGG 780
Db 721 CTACCAGATGCTAGATTTTGGCGCAAGTTTACAGCGCCCTGAGAGCTTCAGAGGACGGG 779
Qy 781 AGGAATCTG 789
Db 780 AGGAATCTG 788

RESULT 14
LOCUS CF746885

DEFINITION UI-M-HB0-cmb-g-24-0-UI.r1 NIH_BMAP_HE0 Mus musculus CDNA clone
IMAGE:30627143 5', mRNA sequence.
ACCESSION CF746885
VERSION CF746885.1 GI:37643228
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORIGIN

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 854)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished [1999] Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. James Lin Unive cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mousefl.html This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

```

FEATURES
  source
Seq primer: pYX-5.
  Location/Qualifiers
    1..854
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:30627143"
      /tissue_type="whole eye"
      /dev_stage="embryo 12.5,13.5,14.5 dpc"
      /lab_host="DH10B (T1 phage resistant)"
      /clone_lib="NIH BMAP HE0"
      /notes="Organ: Eye; Vector: pYX- Asc; Site 1: EcoR I;
      Site 2: Not I; The library was constructed according
      Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,
      1996. Denatured RNA was size fractionated on a 1% agarose
      gel. First strand cDNA synthesis was primed with oligo-dT
      primer containing a Not I site. Double strand cDNA was
      size selected according to mRNA size fraction, ligated
      with EcoR I adaptor, digested with NotI and then cloned
      directionally into pYX-Asc vector. The library tag
      sequence located between the Not I site and the polyA tail
      is TTATTGAAGT. This library was created for the University
      Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
      Developing Mouse Nervous System', supported by National
      Institute of Mental Health (NIMH)."
```

ORIGIN

Query Match	14.0%;	Score 763;	DB 7;	Length 854;
Best Local Similarity	97.4%;	Pred. No. 5.3e-180;		
Matches 835;	Conservative 0;	Mismatches 16;	Indels 6;	Gaps 6;
Qy	852	GTGCTCAGTACTCTTGCATTGGCTTAGAGGAGGAGACTCCAGTGCAGTGCTCGTCCTCAGA	911	
Db	3	GTGTGTCAGTTCTCTTG-AATTGGCTAGAGGAGGAGACTCCAGTGGCAGTGCTCGTCCTCAGA	61	
Qy	912	-CTGGAGATTAAACTCAAGGCCCCAGGAAAGAGCTTTAGACTCTGCAGCGGCCACTTTTCCT	970	
Qy	971	CCTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCAGTAGCTTCAAGCTGGTCTAGT	1030	
Qy	1031	CTCCACCGAGAACTTGACTCTTTCATCCTCTGGGCCCTCTGAGCGTTCAGAGCTTAGGTAA	1090	
Qy	1091	TTTTCAGGTAGTTTCCCTATCCTCTGACCCCTTCTACTCGGACTACTCTTCCCCAGTTAGG	1150	
Qy	1151	GCTGCGCTGTTCAGAGCAGCGCTGGAGTGGCCAGTTTGTGGGTTTCCGAACACTAAACCCC	1210	

Db	362	AGAGAGCTGTAATCTTCTGAAGATGGTGTGTCACCCCTCAGCCGTTGCGGCGAGAGATGC	421
Qy	1271	GGCAACCGCCTCGAGAAAGTGTCGGCTCTCTCTCAGAAAGGCTCGCGGAAATCCACCA	1330
Db	422	GGCAACCGCCTCGAGAAAGTGTCGGCTCTCTCTCAGAAAGGCTCGCGGAAATCCACCA	481
Qy	1331	CTTTCTGATGAAACGGCTAGAAATTCCTCCAGGCTTAACTAAAGGGCAAGAGTTACCCACCCC	1390
Db	482	CTTTCTGATGAAACGGCTAGAAATTCCTCCAGGCTTAACTAAAGGGCAAGAGTTACCCACCCC	541
Qy	1391	TGACCAAGATAATGGCTATCATAGCTTGGAGAGGAAACATAAATCTTCTCGGATGGACCC	1450
Db	542	TGACCAAGATAATGGCTATCATAGCTTGGAGAGGAAACATAAATCTTCTCGGATGGACCC	601
Qy	1451	ACAACATTCGACAGATAAACCCACACAGGCGGTGTCCTCTGTCGACAGAGCGCGGAGCC	1510
Db	602	ACAACATTCGACAGATAAACCCACACAGGCGGTGTCCTCTGTCGACAGAGCGCGGAGCC	661
Qy	1511	CACGTGAAAAAACCCAGAAATGGTGATTCAGAAAGTTTTCACAGAGCCCCCAGGGAAGCAG	1570
Db	662	CACGTGAAAAAACCCAGAAATGGTGATTCAGAAAGTTTTCACAGAGCCCCCAGGGAAGCAG	721
Qy	1571	TCGTGTTTGTGAATACCCGTCGAAAGAAATCTGAAGAGGACACACTAATGCAACTGA	1630
Db	722	TCGTGTTTGTGAATACCCGTCGAAAGAAATCTGAAGAGGACACACTAATGCAACTGA	780
Qy	1631	CCTCTCAGATAGAGGAGAGAGCCTTCTCTGTTTCTACAGACCAAGTTTGTAGCAACAACT	1690
Db	781	CCTCTCAGATAGAGGAGAGAGCCTTCTCTGTTTCTACAGACCAAGTTTGTAGCAACAACT	837
Qy	1691	GATAGATTATATTTTGG	1707
Db	838	GATAGATTATATTTTGG	854
RESULT	15		
LOCUS	CF617044	835 bp	mRNA linear EST 01-OCT-2003
DEFINITION	AGENCOUTT_15763558 NIH MGC_204	Mus musculus	cdna clone
ACCESSION	IMAGE:30527269	5', mRNA	sequence.
VERSION	CF617044		
KEYWORDS	CF617044.1	GI:37233787	
SOURCE	Mus musculus	(house mouse)	
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1. (bases 1 to 835)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Daniela S. Gerhard, Ph.D.		
	Office of Cancer Genomics		
	National Cancer Institute / NIH		
	Bldg. 31 Rm10A07 Bethesda, MD 20892		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Naryan Bhat		
	cdna Library Preparation: Express Genomics		
	cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	Clone sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: NDAM611 row: f column: 14		
	High quality sequence stop: 662.		
FEATURES	Location/Qualifiers		
source	1..835		

FEATURES
SOURCE

/note="Organ: placenta; Vector: pExpress-1; Site: 1: EcoRV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGAGCGCGCCG(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanomolality library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."

Search completed: September 16, 2005, 08:56:22
Job time : 16341.4 secs

ORIGIN

Query Match	13.9%	Score 759.6;	DB 7;	Length 835;
Best Local Similarity	99.2%	Pred. No. 3.8e-179;		
Matches 784;	Conservative 0;	Mismatches 4;	Indels 2;	Gaps 2;
Qy	3992	CAAACTTAACATATTTGGTTTCTGTGCAAACTGTCATTTGGAAGGTTTCTGTGTGTGT 4051		
Db	3	CCAACTTAACATATTTGGTTTCTGTGCAAACTGTCATTTGGAAGGTTTCTGTGTGTGT 62		
Qy	4052	AGTTTTCAGTTCGAATGTCCTCTGTAGCTTTCAAGATGGTAGAATCATTAAGCACT 4111		
Db	63	AGTTTTCAGTTCGAATGTCCTCTGTAGCTTTCAAGATGGTAGAATCATTAAGCACT 122		
Qy	4112	CTTAAGTAATCATTTGCAATTTAGACATTTTTTTTTTTTAACTAGGGGGTATTTGGGAGAC 4171		
Db	123	CTTAAGTAATCATTTGCAATTTAGACATTTTTTTTTTTTAACTAGGGGGTATTTGGGAGAC 181		
Qy	4172	ATGTCAGATATTCATTTTGTGTTTATGTGCTTAAACCAGTGTTACTTACACCCCTGCCT 4231		
Db	182	ATGTCAGATATTCATTTTGTGTTTATGTGCTTAAACCAGTGTTACTTACACCCCTGCCT 241		
Qy	4232	CAGCACTGGCACTTTTCAAACTGCTTGGAGACACTGTAACTTGGATGGTGAGTTCT 4291		
Db	242	CAGCACTGGCACTTTTCAAACTGCTTGGAGACACTGTAACTTGGATGGTGAGTTCT 301		
Qy	4292	GGGTTTTTCATGTGTAATAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTTCTAC 4351		
Db	302	GGGTTTTTCATGTGTAATAATGTCAACTACAAAGGTCAATATCCAGGTTTGTGTGTTCTAC 361		
Qy	4352	CTTATGTAGAGAGAACAAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGCTAGTGT- 4410		
Db	362	CTTATGTAGAGAGAACAAAGCAGAAAGGGCAGATAGAGCAGCCAGAAAGTGCTAGTGT 421		
Qy	4411	CCCCCCCCAAGCGTCTCTAGATAGTGTGGATCAGGTGCTTGGTTGTTGTTCACTTA 4470		
Db	422	CCCCCCCCAAGCGTCTCTAGATAGTGTGGATCAGGTGCTTGGTTGTTGTTCACTTA 481		
Qy	4471	CGGTCACTGTGTGTCATGGTTGAGGTAATGTGTTTACCAGTTCTTTAGTGGTTACATGCA 4530		
Db	482	CGGTCACTGTGTGTCATGGTTGAGGTAATGTGTTTACCAGTTCTTTAGTGGTTACATGCA 541		
Qy	4531	CAAAAGAGAGACCTCTGAGTCGGGTGCGGATGAGCTTTCCAGACCTGGCAGGGTAAAC 4590		
Db	542	CAAAAGAGAGACCTCTGAGTCGGGTGCGGATGAGCTTTCCAGACCTGGCAGGGTAAAC 601		
Qy	4591	TACCTCAGTTTATATCTCCCTGGTTTATTTCCGTTTGTGATGATCTAAGGTCGCTCAG 4650		
Db	602	TACCTCAGTTTATATCTCCCTGGTTTATTTCCGTTTGTGATGATCTAAGGTCGCTCAG 661		
Qy	4651	TGGTGATGATGTTTATCCACACACAAAGGTAGTAAGAGTGCACGACAGAAACGTTGGTC 4710		
Db	662	TGGTGATGATGTTTATCCACACACAAAGGTAGTAAGAGTGCACGACAGAAACGTTGGTC 721		
Qy	4711	TTATTTTGGAGACCCCACTTCTGTGTAATTTATGCACTTGCCTTTTGTAGTGAATCCAGA 4770		
Db	722	TTATTTTGGAGACCCCACTTCTGTGTAATTTATGCACTTGCCTTTTGTAGTGAATCCAGA 781		
Qy	4771	GTGCATTAAA 4780		
Db	782	GTGCATTAAA 791		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2005, 16:13:49 ; Search time 68.7612 Seconds
(without alignments)
3926.035 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHARKRPGRLGSWF.....MFNRLRIESKDLLYSNVKX 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3710	100.0	698	8	AQ88460 Mouse GAD
2	2223	59.9	713	8	AQ88458 Human GAD
3	2213	59.6	713	4	AAB95876 Human pro
4	2157.5	58.2	707	3	AAB53401 Human col
5	547	14.7	153	4	AAW25448 Human pro
6	204	5.5	657	5	ABB57373 Mouse lsc
7	201	5.4	657	2	AAW41102 Rat Myd11
8	191.5	5.2	674	2	AAW99891 Human gro
9	189.5	5.1	674	3	AAW84366 Amino aci
10	189.5	5.1	674	7	AAW84366 Human pro
11	189.5	5.1	674	7	ADW45119 Human pro
12	189.5	5.1	674	7	ADW45119 Human pro
13	186.5	5.0	674	7	ADW45119 Human pro
14	186.5	5.0	674	8	ADW45119 Human pro
15	182.5	4.9	578	2	AAW79958 Human pro
16	182.5	4.9	578	2	AAW79958 Human pro
17	182.5	4.9	578	2	AAW79958 Human pro
18	171.5	4.6	590	2	AAW41104 Human pro
19	149.5	4.0	923	6	ABU54617 Human gao3
20	149	4.0	1365	8	ADQ39691 Human myo
21	148.5	4.0	676	4	ABB62247 Drosophil
22	147.5	4.0	725	4	AAW95238 Human pro
23	147.5	4.0	835	5	ABB05681 Human dif
24	147.5	4.0	835	5	ADW14035 Human NF-
25	147.5	4.0	835	8	ABM80164 Tumour-as

26	144	3.9	867	5	ABP35633	Abp35633 Fungal ZB
27	141	3.8	917	6	ADA13333	Ada13333 Human int
28	141	3.8	1198	6	ADR09843	Adr09843 Human pro
29	139.5	3.8	883	6	ABP72604	Abp72604 Rat brain
30	139.5	3.8	883	6	ADK67782	Adk67782 Rat glyco
31	139.5	3.8	990	5	AAE22762	Aae22762 Rat PGC-3
32	138.5	3.7	1454	8	ABM84777	Abm84777 Human dia
33	138	3.7	1572	5	ABB97562	Abb97562 Novel hum
34	137	3.7	1572	6	ABP98339	Abp98339 Amino aci
35	137	3.7	1088	8	ADQ65665	Adq65665 Novel hum
36	136.5	3.7	883	6	ABP72603	Abp72603 Rat mutan
37	136.5	3.7	883	6	ADK67779	Adk67779 Rat brain
38	136.5	3.7	1234	8	ADQ39694	Adq39694 Human myo
39	136.5	3.7	1366	8	ADQ39692	Adq39692 Human myo
40	136.5	3.7	1411	8	ADQ39693	Adq39693 Human dia
41	136	3.7	826	6	ABM84423	Abm84423 Human dia
42	133.5	3.6	883	7	ADK62581	Adk62581 Rat Prote
43	132.5	3.6	995	8	ADR98899	Adr98899 Lung spec
44	132	3.6	1715	4	ABM80809	Abm80809 Drosophil
45	131.5	3.5	1662	4	AAU32140	Aau32140 Novel hum
46	131.5	3.5	1855	7	ADK73118	Adk73118 Human cel
47	131.5	3.5	1902	7	ADH77759	Adh77759 Human Ala
48	131.5	3.5	3708	7	ABM85423	Abm85423 Human pro
49	131.5	3.5	4127	6	ABR82115	AbR82115 Human ALM
50	130	3.5	1596	5	AAE20271	Aae20271 Human lun
51	129	3.5	817	4	AAU31937	Aau31937 Novel hum
52	129	3.5	930	6	ABB99465	Abb99465 Amino aci
53	127.5	3.4	406	5	ABG93152	Abg93152 S. cerevi
54	127.5	3.4	1003	8	ADK73317	Adk73317 Tobacco 3
55	127.5	3.4	1078	8	ADQ66807	Adq66807 Novel hum
56	127.5	3.4	1281	8	ADR09773	Adr09773 Human pro
57	127.5	3.4	3913	7	ADK40938	Adk40938 Novel hum
58	127.5	3.4	3913	8	ADRI15659	Adri15659 Kinase 51
59	127	3.4	885	7	ADP28591	Adp28591 Neural Ig
60	127	3.4	920	8	ADQ20144	Adq20144 Human sof
61	127	3.4	996	5	AAE22760	Aae22760 Human PGC
62	127	3.4	1002	5	AAE22759	Aae22759 Human PGC
63	127	3.4	1010	7	ADP28580	Adp28580 Neurologi
64	127	3.4	1023	5	AAE22761	Aae22761 Human PGC
65	127	3.4	1028	5	AD128022	Ad128022 ECMCAD pr
66	127	3.4	1028	7	ADP28577	Adp28577 Neural Ig
67	127	3.4	1028	8	ADSI0519	Adsi0519 Human the
68	127	3.4	1056	7	ADD18754	Add18754 Human dis
69	127	3.4	1454	7	ADD93813	Add93813 Human apo
70	127	3.4	1503	2	AAW48845	Aaw48845 Human rec
71	127	3.4	1503	7	ADD93815	Add93815 Human HJ0
72	127	3.4	1503	8	ADP55733	Adp55733 Human PRO
73	126	3.4	237	2	AAW66187	Aaw66187 Human bla
74	126	3.4	817	7	ADE54704	Ade54704 Rat Prote
75	126	3.4	2759	6	AAO16418	Aao16418 Human nuc
76	125.5	3.4	1197	4	AAW95146	Aaw95146 Human pro
77	125.5	3.4	1251	5	AAU98889	Aau98889 Human pro
78	125.5	3.4	1270	5	ABP69291	Abp69291 Human pol
79	125.5	3.4	1316	7	ADF28764	Adf28764 Human deu
80	125	3.4	2781	3	AAW57453	Aaw57453 Human tra
81	125	3.4	2781	8	ADP42724	Adp42724 Human BPT
82	125	3.4	2781	8	ADQ00984	Adq00984 Human hom
83	125	3.4	2781	8	ADQ18653	Adq18653 Human sof
84	125	3.4	2907	3	AAW57452	Aaw57452 Human tra
85	124.5	3.4	1354	4	ABB62359	Abb62359 Drosophil
86	124.5	3.4	1532	4	AAU71956	Aau71956 Human bon
87	124.5	3.4	1780	2	AAW53863	Aaw53863 Human gra
88	124.5	3.4	1780	3	AAW53863	Aaw53863 Human gra
89	124	3.3	605	3	AAW57950	Aaw57950 Human tra
90	124	3.3	640	4	ABG11513	Abg11513 Novel hum
91	124	3.3	1795	4	ABG21018	Abg21018 Novel hum
92	123.5	3.3	582	4	AAW51816	Aaw51816 Murine ap
93	123.5	3.3	718	4	AAW80140	Aaw80140 Human pro
94	123.5	3.3	4274	4	ABG00972	Abg00972 Novel hum
95	123.5	3.3	4386	4	ABG07375	Abg07375 Novel hum
96	123.5	3.3	4397	4	ABG21944	Abg21944 Novel hum
97	123	3.3	1023	7	ADW44438	Adw44438 Human pro
98	123	3.3	1142	2	AAW81546	Aaw81546 Tumour re


```
PT of GADD34L.
PS Disclosure; SEQ ID NO 2; 30pp; English.
XX
XX
CC The present invention relates to a method of screening several test
CC substances for preventing or treating diseases involving oxidative stress
CC such as neuronal ischemia, heart ischemia, renal damage induced by
CC ischemia or toxins, autoimmune diseases and neurodegenerative disorders.
CC The method involves testing the test substances for its ability to
CC inhibit the activity of GADD34-like (GADD34L), also referred to as
CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the
CC test substance which inhibits the activity of GADD34L. the present
CC sequence is the human GADD34L protein.
XX
XX
SQ Sequence 713 AA;

  Query Match      59.9%; Score 2223; DB 8; Length 713;
  Best Local Similarity 65.0%; Pred. No. 5.6e-190;
  Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHARKRPGRLGFWRLPFL-RRSHACSEFFPPSSRONPGN-----SALPERRT 54
Db 1 MEPTGSGRKLGRAGFRFWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60

Qy 55 RYWKLLSQLALLPSLFQKLLWSQLSGGLIPTRWLDFAAYSALRASGRESADPTV 114
Db 61 SYWTKLLSQLALPLGLLQKVLWSQLFGGFWPTRWLDFAGVYSALRALKGREKPAAPTA 120

Qy 115 QKSLSYTAAGLFAKTRVVSTLALARGGTPVAVLRLEVKLKAQERALDSNAPTFILEEQ 174
Db 121 QKSLSSQLQD--SSDPSTVSPLDWLEEGTHWQYSPDLKLEKAGSALDPAQAFLLEEQ 179

Qy 175 LMGVELLPSSIQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSVLYLPOLGRQC 234
Db 180 LMGVELLPSSIQSLRYNRELSSPSPGLNQLQINDFVSVLYLNPSVLYLDFPRLVSYQ 239

Qy 235 SSAGGQFVGPRTTLTPESCYLEDCGHPQPLRAEMSATAMRCCPLSTEGLEPIHHRMR 294
Db 240 NSDGNSEVVGFTLTPESSCLREDHCHPQLSABLIPASWQCPFLSTEGLEPIHHLRMK 299

Qy 295 WLVEL-QPNQODLPTLDQNGYHSLREHNLRWDPOHCTDNPAQVSPAADRP----E 349
Db 300 RLEFLQANKQODLPTPDQNGYHSLREHSLRLMDPKHCRDNPQTQFVPAAGDIPGNTQE 359

Qy 350 PTEKKPELVIOGV-----SQSPQGSLLFCELPVKECEEDHTNATDLSDRGESLPVSTRP 404
Db 360 STEEKIELLTTEVPLALEEESPSCGSPSEIPEKEPEGEGRISVVDYSYLEGDLPIGARP 419

Qy 405 VCSNKLIDYILGGAPSDLEASDSEEDWGEEDDGFDPDGSLSSESDVEODSEGLHLWN 464
Db 420 ACSNKLIDYILGGASSDLETSSDPGEDEWDEAEDDGFDDSSLSLSDGLEQDPGLHLWN 479

Qy 465 SFHSDVPYKPNQFTATTQTAARIAPRDPDSGTGWSGCGV-GSCQGPLPETHDSGE 523
Db 480 SFCSDVPYKPNQFTATTQTAARIIVPEEPESEKDLGSKSDLENSQSGLPETPEHSGE 539

Qy 524 EDDPEPSADEENLKNWSFCHSDPYNLNFKAPFOPSGKNWKGRODSKASSVTVAFS 583
Db 540 EDDMESSADEAESLKNWSFCNSDDPYNLNFKAPFQTSGENEKGCRDSTKTPSEIVAIS 599

Qy 594 GHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCKVQLLSQESCECPDSQVORDVLSGGRHTHVKRKKVTFLEEVTEYIISGDEDRK 659

Qy 644 GPWEEFARDGCRFQKRIQETEVAGYCLAFPHREKMFNRLR 684
Db 660 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLQ 700

RESULT 3
AAB95876
ID AAB95876 standard; protein; 713 AA.
XX
```

```
AC AAB95876;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:18965.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX Claim 8; SEQ ID NO 18965; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX Sequence 713 AA;

  Query Match      59.6%; Score 2213; DB 4; Length 713;
  Best Local Similarity 64.8%; Pred. No. 4.4e-189;
  Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHARKRPGRLGFWRLPFL-RRSHACSEFFPPSSRONPGN-----SALPERRT 54
Db 1 MEPTGSGRKLGRAGFRFWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60

Qy 55 RYWKLLSQLALLPSLFQKLLWSQLSGGLIPTRWLDFAAYSALRASGRESADPTV 114
Db 61 SYWTKLLSQLALPLGLLQKVLWSQLFGGFWPTRWLDFAGVYSALRALKGREKPAAPTA 120
```


Db 606 LGSQSECPDSVQRDVLVSGRHTVKKVTFLEEVTEYYISGDDRKGPWEFPARDGCR 665
Qy 656 FQKRIQTEVAIGYCLAFEHREKMFNRLR 684
Db 666 FQKRIQTEDAIGYCLTFEHRERMFNRLQ 694

RESULT 5
ID AAM25448
AC AAM25448 standard; protein; 153 AA.
XX AAM25448;
XX 16-OCT-2001 (first entry)
DE Human protein sequence SEQ ID NO:963.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antitumagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

XX Homo sapiens.
XX OS
XX W0200153455-A2.
XX 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US035017.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457603/49.
DR N-PSDB; AAM99389.
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX Claim 20; Page 204; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antitumagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders
XX
SQ Sequence 153 AA;
Query Match 14.7%; Score 547; DB 4; Length 153;
Best Local Similarity 76.6%; Pred. No. 1.9e-40;
Matches 105; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

Qy 548 DPNLNFKAPFQPSGKRWKGRDSSKASSEVTVAFSGHTLLSCKAQLLESQDNCEGCG 607
Db 4 DPNLNFKAPFQPSGKRWKGRDSSKASSEVTVAFSGHTLLSCKAQLLESQDNCEGCG 63
Qy 608 LGBALAGERVTHIKRKVTFLEEVTEYYISGDDRKGPWEFPARDGCRFQKRIQETEVAI 667
Db 64 QRDVLSGGRHTVKKVTFLEEVTEYYISGDDRKGPWEFPARDGCRFQKRIQETEDAI 123
Qy 668 GYCLAFEHREKMFNRLR 684
Db 124 GYCLTFEHRERMFNRLQ 140

RESULT 6
ID ABB57373
AC ABB57373 standard; protein; 657 AA.
XX ABB57373;
XX 07-MAR-2002 (first entry)
DE Mouse ischaemic condition related protein sequence SEQ ID NO:1062.
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.
XX Mus musculus.
XX W0200188188-A2.
XX 22-NOV-2001.
XX 18-MAY-2001; 2001WO-JP004192.
XX 18-MAY-2000; 2000JP-00145977.
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX WPI; 2002-034733/04.
DR N-PSDB; ABI99909.
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or by
XX determining the expression profile of a gene group comprising these
XX genes.
XX Claim 2; Page 2676-2679; 2690pp; English.

XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
CC protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or

CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention

XX Sequence 657 AA;

Query Match 5.5%; Score 204; DB 5; Length 657;
 Best Local Similarity 21.5%; Pred. No. 1.2e-08;
 Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

```

QY 60 LLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRRESDAPTQVQSLS 119
DB 119 LLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRRESDAPTQVQSLS 119
QY 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLVLEVKLKAQERAL---DSAPTFL 171
DB 171 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLVLEVKLKAQERAL---DSAPTFL 171
QY 172 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NQKVSYSY---LLN 219
DB 219 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NQKVSYSY---LLN 219
QY 103 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAGLQADKRLGEVVAEEGVAE 160
DB 160 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAGLQADKRLGEVVAEEGVAE 160
QY 220 PSYLDYLPQLGLRCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279
DB 279 PSYLDYLPQLGLRCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279
QY 161 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASIAPGYKPS 199
DB 199 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASIAPGYKPS 199
QY 280 LSTEGLEPIHRRMRWLVPQNGQDLPTLDQNGVHSLSEEHNLRLMDPQHCTDNPAQ 339
DB 339 LSTEGLEPIHRRMRWLVPQNGQDLPTLDQNGVHSLSEEHNLRLMDPQHCTDNPAQ 339
QY 200 TPVPFLGAEHQATE-----EKGTEKADPNSPSSGSHRAWY-YREKPKQGEAKVE 254
DB 254 TPVPFLGAEHQATE-----EKGTEKADPNSPSSGSHRAWY-YREKPKQGEAKVE 254
QY 340 AVSPAADRP-----EPTKEKPELVIOEVSQSPQSSSLFCFLPVEKECEDHTNATDLSDRG 395
DB 395 AVSPAADRP-----EPTKEKPELVIOEVSQSPQSSSLFCFLPVEKECEDHTNATDLSDRG 395
QY 255 AHRAGQGHPCRNAEAEEGGPETTF--VCTGNAFLKAWVYRPGEDTEEDSDSDSAEDT 312
DB 312 AHRAGQGHPCRNAEAEEGGPETTF--VCTGNAFLKAWVYRPGEDTEEDSDSDSAEDT 312
QY 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435
DB 435 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435
QY 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTGTATGTPHTSAFLKAWVYRPG 372
DB 372 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTGTATGTPHTSAFLKAWVYRPG 372
QY 436 EPEDDGFDGSLSDSEVDQSEGLHLWNSFHSVDPYKPNFTATQTAARIAPRDPDS 495
DB 495 EPEDDGFDGSLSDSEVDQSEGLHLWNSFHSVDPYKPNFTATQTAARIAPRDPDS 495
QY 373 DTEENSDDLDS--AEEDTATGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTATQ 429
DB 429 DTEENSDDLDS--AEEDTATGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTATQ 429
QY 496 GTSMGSCGVGSCQBPPLPETPDHSSGDEEDWEPSSADEAKLWNSFCHSDSEDPYNLNF 555
DB 555 GTSMGSCGVGSCQBPPLPETPDHSSGDEEDWEPSSADEAKLWNSFCHSDSEDPYNLNF 555
QY 430 GATPHTSPFLKAWVYRPGEDTEEDSDSENVAPGDSADSSQSPC-----478
DB 478 GATPHTSPFLKAWVYRPGEDTEEDSDSENVAPGDSADSSQSPC-----478
QY 556 KAPFOPS-----GKNWKGQDSKASSEVTVFSGHH-----TLSCKAQLLES--598
DB 598 KAPFOPS-----GKNWKGQDSKASSEVTVFSGHH-----TLSCKAQLLES--598
QY 479 ---LQPRCLPGEKTKGRGEPPLFQVAFYLPGEKPSFWAAPKLPLRLQRLRLFKAPT 535
DB 535 ---LQPRCLPGEKTKGRGEPPLFQVAFYLPGEKPSFWAAPKLPLRLQRLRLFKAPT 535
QY 599 --QEDNCPGCGLGEALAGERVTHIKKVKVTFLEEVTEYIIS-----GDEDRKGWEEFAR 651
DB 651 --QEDNCPGCGLGEALAGERVTHIKKVKVTFLEEVTEYIIS-----GDEDRKGWEEFAR 651
QY 536 RDQDEIP-----LKAHVHAEKVTVHFLAWAGPAQAARRGWEEQAR 580
DB 580 RDQDEIP-----LKAHVHAEKVTVHFLAWAGPAQAARRGWEEQAR 580
QY 652 DGCPRKRIQETEAIGYCLAPEHREKMFNRLR 684
DB 684 DGCPRKRIQETEAIGYCLAPEHREKMFNRLR 684
QY 591 DRSRPARIAQAEKLGPLYLPDSRARAWARL 613
DB 613 DRSRPARIAQAEKLGPLYLPDSRARAWARL 613

```

RESULT 7

AA41102

ID AA41102 standard; protein; 657 AA.

AC AA41102;

XX 17-JAN-2000 (first entry)

XX Rat MyD116 protein sequence.

XX Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;

XX cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;

XX MyD116 protein.

OS Rattus sp.

XX WO9949898-A1.

XX 07-OCT-1999.

XX 31-MAR-1999; 99WO-US007199.

XX 31-MAR-1998; 98US-00052753.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Fisher PB;

XX WPI; 1999-591184/50.

XX Novel vectors useful for studying the progression of cancer.

XX Disclosure; Fig 3A-B; 251pp; English.

XX The invention relates to an inducible progression-elevated gene-3 (PEG-3

XX gene) regulatory region functionally linked to a gene encoding a product

XX that causes or may be induced to cause the death or inhibition of cancer

XX cell growth. A vector of the invention which contains a gene encoding

XX thymidine kinase or a product which causes the cell to express a specific

XX antigen can be administered along with gancyclovir or acyclovir, or an

XX antibody or fragment to the antigen, respectively, to treat cancer in a

XX subject. The PEG-3 gene is useful for generating new cloning and

XX expression vectors, transfected cells, and for developing methods for

XX cultured growth of such cells. The PEG-3 polynucleotide is also useful as

XX a source of primers and probes to study the progression of cancer, and to

XX detect the presence of the gene. The present sequence represents the

XX amino acid sequence of the rat MyD116 protein

XX Sequence 657 AA;

Query Match 5.4%; Score 201; DB 2; Length 657;

Best Local Similarity 21.5%; Pred. No. 2.3e-08;

Matches 149; Conservative 85; Mismatches 291; Indels 168; Gaps 27;

QY 60 LLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRRESDAPTQVQSLS 119

DB 119 LLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRRESDAPTQVQSLS 119

QY 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLVLEVKLKAQERAL---DSAPTFL 171

DB 171 YTAAGLFAKTRVSTLALARG-----GTPVAVLVLRLVLEVKLKAQERAL---DSAPTFL 171

QY 172 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NQKVSYSY---LLN 219

DB 219 EQQLMGVELLPS-SLQAGLVSHRELD-SSSGPLSVQSLG-----NQKVSYSY---LLN 219

QY 103 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAGLQADKRLGEVVAEEGVAE 160

DB 160 --ETWGLSNVDYNAKPGQDDLREKEMERTAGKATLQAGLQADKRLGEVVAEEGVAE 160

QY 220 PSYLDYLPQLGLRCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279

DB 279 PSYLDYLPQLGLRCQSSAGGQFVGFRYLTLPESCYSLEDGCHPQLRAEMSATWRRCPP 279

QY 161 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASIAPGYKPS 199

DB 199 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASIAPGYKPS 199

QY 280 LSTEGLEPIHRRMRWLVPQNGQDLPTLDQNGVHSLSEEHNLRLMDPQHCTDNPAQ 339

DB 339 LSTEGLEPIHRRMRWLVPQNGQDLPTLDQNGVHSLSEEHNLRLMDPQHCTDNPAQ 339

QY 200 TPVPFLGAEHQATE-----EKGTEKADPNSPSSGSHRAWY-YREKPKQGEAKVE 254

DB 254 TPVPFLGAEHQATE-----EKGTEKADPNSPSSGSHRAWY-YREKPKQGEAKVE 254

QY 340 AVSPAADRP-----EPTKEKPELVIOEVSQSPQSSSLFCFLPVEKECEDHTNATDLSDRG 395

DB 395 AVSPAADRP-----EPTKEKPELVIOEVSQSPQSSSLFCFLPVEKECEDHTNATDLSDRG 395

QY 255 AHRAGQGHPCRNAEAEEGGPETTF--VCTGNAFLKAWVYRPGEDTEEDSDSDSAEDT 312

DB 312 AHRAGQGHPCRNAEAEEGGPETTF--VCTGNAFLKAWVYRPGEDTEEDSDSDSAEDT 312

QY 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435

DB 435 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSESED-----W-----GE 435

QY 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTGTATGTPHTSAFLKAWVYRPG 372

DB 372 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTGTATGTPHTSAFLKAWVYRPG 372

QY 436 EPEDDGFDGSLSDSEVDQSEGLHLWNSFHSVDPYKPNFTATQTAARIAPRDPDS 495

DB 495 EPEDDGFDGSLSDSEVDQSEGLHLWNSFHSVDPYKPNFTATQTAARIAPRDPDS 495

QY 373 DTEENSDDLDS--AEEDTATGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTATQ 429

DB 429 DTEENSDDLDS--AEEDTATGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTATQ 429

QY 496 GTSMGSCGVGSCQBPPLPETPDHSSGDEEDWEPSSADEAKLWNSFCHSDSEDPYNLNF 555

DB 555 GTSMGSCGVGSCQBPPLPETPDHSSGDEEDWEPSSADEAKLWNSFCHSDSEDPYNLNF 555


```

Db 430 GATPHTSPFLKAWYRFGEDTDDTEEDSENAPGDSETADSSQSPC----- 478
Qy 556 KAPFQPS---GKMKGRQDSKASSEVTVAFSGH-----TLLSKAQLLS--- 598
Db 479 ---LQPORCLPGEKTKGRGEPPLFQVAFYLPGEKPSWAAPKLPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGLGBALAGERYTHIKKVKVTFLEEVYIYS-----GDEDRKGWEEFAR 651
Db 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGWEQFAR 580
Qy 652 DGCRFQRIQETVAIGCYCLAFEHREKMFNRLR 684
Db 581 DRSRFAARIAQAEKGLGYLTDPDSRARAWARLR 613

RESULT 8
AAW99891
ID AAW99891 standard; protein; 674 AA.
AC AAW99891;
DT 09-JUN-1999 (first entry)
DE Human growth regulator protein GRREG.
KW Human; growth regulator protein; GRREG; cancer.
OS Homo sapiens.
XX WO9902680-A1.
XX 21-JAN-1999.
XX 30-JUN-1998; 98WO-US013409.
XX 11-JUL-1997; 97US-00893852.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Lal P, Shah P, Corley NC;
XX WPI; 1999-120881/10.
XX N-PSDB; AAX19798.
XX Novel human growth regulator protein - useful in the treatment of
XX cancers.
XX Claim 1; Fig 1; 70pp; English.
XX The present sequence represents a human growth regulator protein,
XX designated GRREG. The protein can be used in methods to treat cancer. The
XX protein can also be used in binding assays to detect (ant)agonists of
XX GRREG. A fragment from the nucleic acid sequence encoding GRREG can be
XX used as a probe for detecting GRREG encoding sequences (especially in PCR
XX amplified samples)
XX Sequence 674 AA;

Query Match          5.2%; Score 191.5; DB 2; Length 674;
Best Local Similarity 21.5%; Pred. No. 1.7e-07;
Matches 156; Conservative 75; Mismatches 268; Indels 227; Gaps 32;

Qy 60 LLSQALLPSLFQKLLWSQLGLIPTR-WLDPFASYSALRASRGREESDPTVQKSL 118
Db 21 LLSVMSLLSR-----AWSRLR-GLGLEPWLVEAVKGAAL----- 55
Qy 119 SYTAAGLFAKTRVVTALARGGTFVAVLVLRLVLEVKLKAQERALDSAAP-----T 168
Db 56 --VQAGLEGEAR-----TPLA--IPHTPWGRPRPEEAEDSGPGEDRETGLKT 100
Qy 169 FLLEQQWQVVELLPSSLOAGLVSHRELDSSSSGSLSVQSLGNFKVSYLLNPSYL----- 223

```

```

Db 101 SSSLPEAWG--LLDD--DDMGYGEREATSVPRQGSQFADGQ-----RAPLSPSLLIRTQ 152
Qy 224 --DYLQGLRQCSSAGGQFVGFRLT-----PESCYLESDGCHPQPLRAEMSATA--- 273
Db 153 GSDKNP--GEEKAEEGVAEEEGVNFSPSHRECCPAVEEEDDEEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPEIHHRRWRLVFIQPNQOGDLPTLDQONGVHSL 320
Db 211 LSPGSKPSTWVSCPGSEENQATBDKRTSKGARKTSVSPRSSGSDPRSWEYRSGEASE 270
Qy 321 EENLLRMPDQHCNDNPAQAVSPAADRPEKPELVQEVQSVQSGSSLPFCLEPVEKE 380
Db 271 KE-----EKAHEETKGEAA-----PGQSSAP-----AQRQLKSNWQCPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRFVCSNKLIDYILGAPSDLEASSDSESEDWGEPEDD 440
Db 312 SEVKALGAAEKGEAEACPPCIPPSAFLKAWVWPG--EDTEEEDEEB-----DESDS 364
Qy 441 GFDSDGSLSESDVEQDSEGLHL--WNSFHSVDYKQNFNTATIQ-----TAAARIAPRDP 493
Db 365 GSDEEEGEAEASSSTPATGVFLKSW-----VYQPGEDTEEEDESDTGSADEREAE 417
Qy 494 DSCTSWSGCGVSCQEGPLPETPDH-----SSGEEDDWEPADAEALNKLWNSPCHSE 547
Db 418 TSASTPPASAFKAWYRPGEDTEEEDEEDVDSEKDDSEALGAEAS----- 466
Qy 548 DYNLLNFKAQFQPSG---KNWKGQDSKASSEVTVAFSGHHTLLSCKAQLLESQDN 603
Db 467 DPH-----PSHPDOSAHFRGHWYRPGKETEE-----EAEED-- 498
Qy 604 PGCLGEA-----LAGER-----YTHIKRK 623
Db 499 ---WGEABPCPPRVAIVYVGEKPPPPWAPPLRLRLQRLKRPEPTHTDPPETPLKAR 554
Qy 624 KYTFLEEVTEYIYS-----GDEDRKGPWEFADGCRFQKRIQETVAIGCYCLAFEHREK 678
Db 555 KVRPSEKVTVHFLAVWAGPAQAARQGWELQALDRSRFARRITQAEELSCLTTPAAR 614
Qy 679 MFNRLR 684
Db 615 AWARLR 620

RESULT 9
AAV84366
ID AAV84366 standard; protein; 674 AA.
XX AAV84366;
XX 12-JUL-2000 (first entry)
DT Amino acid sequence of a human apoptosis associated protein (GADD34).
DE Human; apoptosis associated protein; GADD34; antiviral agent;
XX growth arrest and DNA damage-inducible gene 34; viral infection;
KW serine/threonine phosphatase.
XX Homo sapiens.
XX GB2342716-A.
XX 19-APR-2000.
XX 27-JUL-1999; 99GB-00017631.
XX 14-SEP-1998; 98GB-00020025.
XX (PFIZ ) PFIZER LTD.
XX Ciaramella G;
XX WPI; 2000-285397/25.
XX N-PSDB; AAZ99766.

```

XX	Screening assay for potential antiviral agents that modify serine/threonine phosphatase activity for treating human viral infections.	Db	555	KVRFSEKVTYHFLAVWAGPAQAARQGPWEQIARDRSFARRITQAEBSPLTPAARAR	614
PT		Qy	679	MFNELR 684	
PT		Db	615	AWARLR 620	
XX	Claim 6; Page 99; 106pp; English.				
CC	The present sequence represents a human apoptosis associated protein (GADD34). It is encoded by a growth arrest and DNA damage-inducible gene 34. It is used in an assay method for identifying an anti-viral agent that can affect the activity or expression of a nucleotide sequence or its expression product. The assay method comprises contacting an agent with a nucleotide sequence coding for a serine/threonine phosphatase or its expression product, and determining whether the agent affects the activity or expression of the nucleotide sequence. The assay is used for screening for antiviral agents useful in the treatment of human viral infections. The anti-viral modulators of type 1 protein phosphatase activity inhibit the action of a host component which interferes with the activation of a host-cell interferon-induced, double-stranded RNA-activated protein kinase e.g. p68, eIF2 or eIF2alpha				
XX	Sequence 674 AA;				
Query Match	5.1%; Score 189.5; DB 3; Length 674;				
Best Local Similarity	21.5%; Pred. No. 2.6e-07;				
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;					
Qy	60	LLSQLALLPSLFQKLLWSQLSGGLIPTR-WLDPAAASYALRASRGRESDAPTVQKSL	118		
Db	21	LLSPVMGLSR-----AWSRLR-GIGPLEPWLVEAVKGAAL-----	55		
Qy	119	SYTAAGLPAKTRVNSTALARGGTPVAVLVLRLVKLKAQERALDSAAP-----T	168		
Db	56	--VEAGLSEAR-----IPLA--IHTPWGRPRPEEAEDSGGPGEDRETGLTKT	100		
Qy	169	FLLEQLMGVLLPSSLAQLVSHLELDSSSGPLSVQSLGNFKVSVLLNPSYL-----	223		
Db	101	SSSLPEAWG--LLDD--DCMGYGEREATSVPRGQSQPADGQ----RAPLSFSLIRTLQ	152		
Qy	224	--DYLQPLGLRCQSSAGGQGVGFRTLT-----PESCVLSEDCGHCPQLRAEMSATA--	273		
Db	153	GSDKNP--GEKAELEGVAEEGVNKFYPSHRECCPAVEEDDEEAIVKKEAHTSTSA	210		
Qy	274	-----WRRCP-----PLSTEGLEPIHRRMRWLVLPNQGGQLFTLDDQNGYHSLE	320		
Db	211	LSPGSKPTWVSCPEEENQATEDKRTSRSGARKTSVSPRSSGSDPRSWEYRSGEASEE	270		
Qy	321	BEHLLRMDOHCTDNPAQVSPAADRPEPEKPELVIEVQSQPQSGSLFCLPLVKE	380		
Db	271	KE-----EKAKHETGKEAA-----PGQSSAP-----AQRPLKSWWCQPSDEEE	311		
Qy	381	CEEDHTNATDLSRGESLPVSTRPVCSNKLIDYILGGAPSDLEASSDSESDWGEPEDD	440		
Db	312	GEVALGAAEKDGEACPCPIPPSPAFUKAWVWPG-----EDTEEEDEEE-----DESDS	364		
Qy	441	GFDSDGSLSESDVEQDSGLHL--WNSPHSVDPKPNFTATIQ-----TAARIAPRDP	493		
Db	365	GSDEEGEAASSTPATGVFLKSW-----VTPQGEDTEEEDESDTGSADEREAE	417		
Qy	494	DSGTSWSGCGVGCQEGPLPETPDH-----SGEEDDWPSPADEANLKLWNSPCHSE	547		
Db	418	TSASTPPASAFKAWVYRPGEDTEEEDEDDVEDKEDDSAAALGEAFS-----	466		
Qy	548	DPYNLLNFKAPQPSG-----KNWGRQDSKASSETVYAFSGHHTLLSCAKQLLESQDNC	603		
Db	467	DPH-----PSHPDQRAHFGWVYRPGKETEE-----FAEED--	498		
Qy	604	PGCGLGEA-----LAGER-----YTHIKRK	623		
Db	499	---WGEAECPEFVAIVVPEKPPPPWAPRLPLRLQRLKRPTETPHDPPETPLKAR	554		
Qy	624	KVTFLEEVTEYIS-----GDEDRKGWBEFARCGCRFKRIQETEAIVGCLAFHEK	678		

Db	555	KVRFSEKVTYHFLAVWAGPAQAARQGPWEQIARDRSFARRITQAEBSPLTPAARAR	614
Qy	679	MFNELR 684	
Db	615	AWARLR 620	
RESULT 10			
ADE54748			
ID	ADE54748	standard; protein; 674 AA.	
XX	AC	ADE54748;	
XX	DT	29-JAN-2004 (first entry)	
XX	DE	Human Protein XP_009097, SEQ ID NO 553.	
XX	KW	Human; pain; neuronal tissue; gene therapy;	
KW	KW	spinal segmental nerve injury; chronic constriction injury; CCI;	
KW	KW	spared nerve injury; SNI; Chung.	
XX	OS	Homo sapiens.	
XX	FN	WO2003016475-A2.	
XX	PD	27-FEB-2003.	
XX	PF	14-AUG-2002; 2002WO-US025765.	
XX	PR	14-AUG-2001; 2001US-0312147P.	
PR	PR	01-NOV-2001; 2001US-0346382P.	
PR	PR	26-NOV-2001; 2001US-0333347P.	
XX	PA	(GPHO) GEN HOSPITAL CORP.	
PA	PA	(FARB) BAYER AG.	
XX	PI	Woelf C, D'urso D, Befort K, Costigan M;	
XX	DR	WPI; 2003-268312/26.	
DR	DR	GENBANK; XP_009097.	
XX	PT	New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.	
XX	PS	Claim 1; Page; 1017pp; English.	
CC	CC	The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	

XX	Sequence 674 AA;	
SO	Query Match	5.1%; Score 189.5; DB 7; Length 674;
	Best Local Similarity	21.5%; Pred. No. 2.6e-07;
	Mismatches	156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;
Qy	60	LLSQALLPSLFQKLLWSQSGGLIPTR-WLDPFAASYSALRASRGRESDAPTQVQSL 118
Db	21	LLSPVMGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy	119	SYTAAGLFAKTRVVTALARGGTPVAVLRLVLEVKLKAQERALDSAP-----T 168
Db	56	--VEAGLEGEAR-----TPLA--IPHTPWGRRPEEAEADSGGCGEDRETGLKT 100
Qy	169	FLLEQQLWGVLELPSLSLQAGLVSHRELDSSSSGSLSVQSLGNPKVSVYLLNPSYL----- 223
Db	101	SSSLPEAWG--LLDD--DDGMYGEREATSVPRGQSQFADGQ-----RAPLSPLLRTLQ 152
Qy	224	--DYLQGLRQSSAGGQFVGFRILT-----PESCYLSEDCGCHPOPLRAEMSATA--- 273
XX	Sequence 674 AA;	
OS	Query Match	5.1%; Score 189.5; DB 7; Length 674;
PN	Best Local Similarity	21.5%; Pred. No. 2.6e-07;
XX	Mismatches	156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;
PD	27-FEB-2003.	
XX	14-AUG-2002; 2002WO-US025765.	
PF	14-AUG-2001; 2001US-0312147P.	
PR	01-NOV-2001; 2001US-0346382P.	
PR	26-NOV-2001; 2001US-0333347P.	
XX	(GEOH) GEN HOSPITAL CORP.	
PA	(FARB) BAYER AG.	
XX	Woelf C, D'urso D, Befort K, Costigan M;	
XX	WPI; 2003-268312/26.	
DR	GENBANK; XP_009097.	
XX	New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.	
PT	Claim 1; Page; 1017pp; English.	
PT	The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 674 AA;	
Qy	60	LLSQALLPSLFQKLLWSQSGGLIPTR-WLDPFAASYSALRASRGRESDAPTQVQSL 118
Db	21	LLSPVMGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy	119	SYTAAGLFAKTRVVTALARGGTPVAVLRLVLEVKLKAQERALDSAP-----T 168
Db	56	--VEAGLEGEAR-----TPLA--IPHTPWGRRPEEAEADSGGCGEDRETGLKT 100
Qy	169	FLLEQQLWGVLELPSLSLQAGLVSHRELDSSSSGSLSVQSLGNPKVSVYLLNPSYL----- 223
Db	101	SSSLPEAWG--LLDD--DDGMYGEREATSVPRGQSQFADGQ-----RAPLSPLLRTLQ 152
Qy	224	--DYLQGLRQSSAGGQFVGFRILT-----PESCYLSEDCGCHPOPLRAEMSATA--- 273

XX	Sequence 674 AA;	
SO	Query Match	5.1%; Score 189.5; DB 7; Length 674;
PN	Best Local Similarity	21.5%; Pred. No. 2.6e-07;
XX	Mismatches	156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;
PD	27-FEB-2003.	
XX	14-AUG-2002; 2002WO-US025765.	
PF	14-AUG-2001; 2001US-0312147P.	
PR	01-NOV-2001; 2001US-0346382P.	
PR	26-NOV-2001; 2001US-0333347P.	
XX	(GEOH) GEN HOSPITAL CORP.	
PA	(FARB) BAYER AG.	
XX	Woelf C, D'urso D, Befort K, Costigan M;	
XX	WPI; 2003-268312/26.	
DR	GENBANK; XP_009097.	
XX	New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.	
PT	Claim 1; Page; 1017pp; English.	
PT	The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 674 AA;	
Qy	60	LLSQALLPSLFQKLLWSQSGGLIPTR-WLDPFAASYSALRASRGRESDAPTQVQSL 118
Db	21	LLSPVMGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy	119	SYTAAGLFAKTRVVTALARGGTPVAVLRLVLEVKLKAQERALDSAP-----T 168
Db	56	--VEAGLEGEAR-----TPLA--IPHTPWGRRPEEAEADSGGCGEDRETGLKT 100
Qy	169	FLLEQQLWGVLELPSLSLQAGLVSHRELDSSSSGSLSVQSLGNPKVSVYLLNPSYL----- 223
Db	101	SSSLPEAWG--LLDD--DDGMYGEREATSVPRGQSQFADGQ-----RAPLSPLLRTLQ 152
Qy	224	--DYLQGLRQSSAGGQFVGFRILT-----PESCYLSEDCGCHPOPLRAEMSATA--- 273
Db	153	GSDKNP--GEEKAEVGAEEGVNKFSPSHRECCPAVEEEDDEAVKKAHRTSTA 210
Qy	274	-----WRCRP-----PLSTEGLPPIHRRMRWLFLQPNQOQLPTLDQNGYHSL 320
Db	211	LSPGKPESTWVSCFGEENQATDKRTSKGAKTSVSPRSGSDPRSWEYRSGEASE 270
Qy	321	BEHLLRMDPOHCTDNPAQAVSPAADRPETPKKPELVIOEVSQSGSLFCBEPVEKE 380
Db	271	KE-----EKAKETGGEAA-----PGQSSAP-----AQPQLKSWQPSDEE 311
Qy	381	CEEDHTNATDLSRGSPLVSTRVCSNKLIDYILGAPSDLEASSSESDGEEPEDD 440
Db	312	GEVKALGAERKDGAECPPIPPPSAFKAWVWPG--EDTEEEDEEB-----DESDS 364
Qy	441	GFSDGSLSDVDQDSEGLHL--WNSFHSVDPKPQNFATIQ-----TAARIAPRDS 493
Db	365	GSDEEGEAEASSSTPATGVFLKSW-----VYQGEDTEEEDESDTGSADEREAE 417
Qy	494	DSGTSWSGSCGVGSCQGLPETPDH-----SSGEEDDWEPSADAEENLKNWFSCHSE 547
Db	418	TSASTPPASAPLKAWVRPBDTEEBEDVDSDKEDDSEALGEABS----- 466
Qy	548	DPYNLLNFKAPFQPSG-----KNWGRQDSKASSETVTVAFSGHHTLLSCKAQLLESQDNC 603
Db	467	DPH-----PSHDQRAHFRGMYRPGKETEER-----EAEED-- 498
Qy	604	PGCLGEA-----LAGER-----YTHIKRK 623
Db	499	-----WGEAEPCEPRVAIVYFGEKPPPPWAPPRLRLQLRRLKRPETPTDPTPLKAR 554
Qy	624	KVTFLEBVTYIIS-----GDENRKGWBERFARDGCRFQKRIQTEVAIGVCLAFAREK 678
Db	555	KVRFSEKVTYHFLVWAGPAQAARQGWEPQEARDRSRFARRITQAQBEELSCITPAARAR 614
Qy	679	MFNRLR 684
Db	615	AWAELR 620
XX	RESULT 11	
ID	ADD45119	
XX	ADD45119 standard; protein; 674 AA.	
XX	AC	ADD45119;
XX	XX	29-JAN-2004 (first entry)
DT	XX	Human Protein XP_009097, SEQ ID NO 10552.
DE	XX	Human; pain; neuronal tissue; gene therapy;
KW	XX	spinal segmental nerve injury; chronic constriction injury; CCI;
KW	XX	spared nerve injury; SNI; Chung.

```
Db 153 GSDKNP--GEEKAEEGVAEEGVNKFSPSPSHRECCPAVEEEDDEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPEIHRRMRWLVPLOQNOQDLPTLDQNGYHSLE 320
Db 211 LSPGSKPSTWVSCFGEENQATDKRTSRGARKTSVSPRSSGSDPRSMWEYRSGEASEE 270
Qy 321 EEHLLRMDPOHCTDNDPAQAVSPAARDPEPTEKKPELVIQVVSQSPGSSSLFCBLPYEKE 380
Db 271 KE-----EKAHKTGKEAA-----PGOSSAP-----AQRPLKSWWCQPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGAPSDLEASSDSESDMGEPEDD 440
Db 312 GEVKALGAEKDGEAECCPCCIPPSAPFLKAWVWPG---EDTEEEDEEE---DESDS 364
Qy 441 GFDSGSLSDVDQDSEGLHL--WNSFHSVDPKPQNFTATIQ-----TAARIAPRDP 493
Db 365 GSDEEGEAESSTPATGVFLKSW-----VYQGEDTEEEDESDTGSADEREAE 417
Qy 494 DSGTSWSGCGVSCQEGPLPETPDH-----SSGEDDWEPSADEAENLKLWNSFCHSE 547
Db 418 TSASTPPASAPFLKAWVYRPGEDTEEEDEDVDSEKEDDSEALGEAES-----466
Qy 548 DPYNLLNFKAPFQSG-----KNWGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603
Db 467 DPH-----PSHPDQRAHFRGWYRPGKETEE-----EAED--498
Qy 604 PGCGLGEA-----LAGER-----YTHIKRK 623
Db 499 -----WGEABCPFRVAYVPGKEPPPPWAPRLPLRLQRLKRPETPHDPDPETPLKAR 554
Qy 624 KVTLEEVTEYIIS-----GBEDRKGPMEEFARDGCRFOKRIQTEVAIGYCLAFEHREK 678
Db 555 KVFSEKTVHFLAVWAGPAQAARQGPWEQLARDSRFARITQAEBSLCLTPAAR 614
Qy 679 MNRLR 684
Db 615 AWALR 620

RESULT 12
ADE54744
ID ADE54744 standard; protein; 674 AA.
XX AC ADE54744;
XX DT 29-JAN-2004 (first entry)
XX DE Human Protein XP_009097, SEQ ID NO 549.
XX KW Human; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GSHO ) GEN HOSPITAL CORP.
XX PA (FARB ) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX DR GENBANK; XP_009097.
```

```
XX New composition comprising two or more isolated polypeptides, useful for
Pt preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
Cc or human polynucleotides or a polynucleotide which represents a fragment,
Cc derivative or allelic variation of the nucleic acid sequence. Also
Cc claimed are a vector comprising the novel polynucleotide, a host cell
Cc comprising the vector, a method for identifying a nucleotide sequence
Cc which is differentially regulated in an animal subjected to pain and a
Cc kit to perform the method, an array, a method for identifying an agent
Cc that increases or decreases the expression of the polynucleotide sequence
Cc that is differentially expressed in neuronal tissue of a first animal
Cc subjected to pain, a method for identifying a compound which regulates
Cc the expression of a polynucleotide sequence which is differentially
Cc expressed in an animal subjected to pain, a method for identifying a
Cc compound that regulates the activity of one or more of the
Cc polynucleotides, a method for producing a pharmaceutical composition, a
Cc method for identifying a compound or small molecule that regulates the
Cc activity in an animal of one or more of the polypeptides given in the
Cc specification, a method for identifying a compound useful in treating
Cc pain and a pharmaceutical composition comprising the one or more
Cc polypeptides or their antibodies. The polynucleotide or the compound that
Cc modulates its activity is useful for preparing a medicament for treating
Cc pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
Cc injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
Cc therapy). The sequence presented is a human protein (shown in Table 2 of
Cc the specification) which is differentially expressed during pain. Note:
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic form directly from WIPO at
Cc ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 674 AA;

Query Match 5.1%; Score 189.5; DB 7; Length 674;
Best Local Similarity 21.5%; Pred. No. 2.6e-07;
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;

Qy 60 LLSQLLALLPSLQKLLNSQLSGGLIPTR-WLDFAASTYSALRASRGRESDAPTQKSL 118
Db 21 LLSFVNGLLSR-----AWSRLR-GLGFLEPWLVEAVKGAAL-----55
Qy 119 SYTAAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAQERALDSAAP-----T 168
Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRPEEAEADSGGREDRETLGKT 100
Qy 169 FLLEQQLMGVELLPSSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLNPVYL-----223
Db 101 SSSLPEAWG--LLDD--DDGMGEREATSVPRGGSQFADGQ----RAPLSPLLIRTLQ 152
Qy 224 --DYLPLQLGLRCOSSAGGQFVGRITL-----PESCYLSEDCGCHPQPLAENSATA---273
Db 153 GSDKNP--GEEKAEEGVAEEGVNKFSPSPSHRECCPAVEEEDDEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPEIHRRMRWLVPLOQNOQDLPTLDQNGYHSLE 320
Db 211 LSPGSKPSTWVSCFGEENQATDKRTSRGARKTSVSPRSSGSDPRSMWEYRSGEASEE 270
Qy 321 EEHLLRMDPOHCTDNDPAQAVSPAARDPEPTEKKPELVIQVVSQSPGSSSLFCBLPYEKE 380
Db 271 KE-----EKAHKTGKEAA-----PGOSSAP-----AQRPLKSWWCQPSDEEE 311
Qy 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGAPSDLEASSDSESDMGEPEDD 440
Db 312 GEVKALGAEKDGEAECCPCCIPPSAPFLKAWVWPG---EDTEEEDEEE---DESDS 364
Qy 441 GFDSGSLSDVDQDSEGLHL--WNSFHSVDPKPQNFTATIQ-----TAARIAPRDP 493
Db 365 GSDEEGEAESSTPATGVFLKSW-----VYQGEDTEEEDESDTGSADEREAE 417
Qy 494 DSGTSWSGCGVSCQEGPLPETPDH-----SSGEDDWEPSADEAENLKLWNSFCHSE 547
```

Db 418 TSASTPPASAFKAWVYRPGDTEEBEDDVSDEKDDSEALGEAES----- 466
 Qy 548 DPYNLLNFKAPFQPSG-----KNWKGQDSKASSVTVAFSGHHTLLSCKAQLLESQDNC 603
 Db 467 DPH-----PSHPDQRAHFRGMYRPGKETEE-----EAEED-- 498
 Qy 604 PGCGLGEA-----LAGER-----YTHIKRK 623
 Db 499 ----WGEAEPFPRVAIYVPGKPPPPWAPPLRLQLRLLKRPETPTDPPETPLKAR 554
 Qy 624 KVTFLLEVTEYIIS-----GDEDRKGPWEFARDGCRFQKRIQETEAIGYCLAFHREK 678
 Db 555 KVRSEKVTYVHFLAVWAGPAQAARQGWELQARDRSFARRITQAQBELSPCLTPAAR 614
 Qy 679 MFNRLR 684
 Db 615 AWARLR 620

RESULT 13
 AAB92888
 ID AAB92888 standard; protein; 674 AA.
 XX
 AC AAB92888;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11494.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ichii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX
 PS Claim 8; SEQ ID NO 11494; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 CC
 SQ Sequence 674 AA;
 Query Match 5.0%; Score 186.5; DB 4; Length 674;
 Best Local Similarity 21.5%; Pred. No. 4.9e-07;
 Matches 156; Conservative 73; Mismatches 270; Indels 227; Gaps 32;
 Qy 60 LLSQLLALLPSLFQKLLMSQLSGGLIPTR-WLDFAASYSALRASRGRESDAPTVQKSL 118
 Db 21 LLSFVWGLLSR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
 Qy 119 SYTAAGLFAKTRVYSTLALARGGTPVAVLVLRLEVKLKAQERALDSAAP-----T 168
 Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRRPGCEAEDSGGPGEDRETGLKT 100
 Qy 169 FLLEQQLWGVLLPSSLOAGLVSHRELDSSSSSPLSVQSLGNPKVSYLLNPSYL----- 223
 Db 101 SSSLPEAWG--LLDD--DDGMGEREATSVPRGQSGQFADGQ----RAPLSPSLLIRTLO 152
 Qy 224 --DYLQGLRCQSSAGGQFVGFRILT-----PESCYLSEDCGCHPOPLRAENSATA-- 273
 Db 153 GSDKNP--GEEKAEERGVAEEGVNKFSPPSHRECCPAVEEEDDEAVKEAHRITSTA 210
 Qy 274 -----WRRCP-----PLSTEGPLPEIHHRRMRVLVFLQPNQOQDLPTLDQDNGYHSL 320
 Db 211 LSPGSKPSTWVSCPGEEENQATEDKRTSKGARKTSVSPRSSGSDPSRWSYRSGEASEE 270
 Qy 321 EBNLLRMDPHOCTDNPAQAVSPAARDPEPTEKKPELVLTQEVVSQSGSLFELPEVKE 380
 Db 271 KE-----EKAHEETGKEAA-----PGQSSAP-----AORPLKSWWCQPSDEE 311
 Qy 381 CEEDHTNATDLSDRGSLPVSTRPVCSNKLDIYLGGAPSDLEASDSESDNGERPEDD 440
 Db 312 SEVKPLGAAEKQGEAECPPCIPPPSAFLKAWVTWPG---EDTEEEDEEE---DESDS 364
 Qy 441 GFDSGSLSESVDQSEGLHL--MNSFHSVDYPKQNFETATIQ-----TAARIAPRDPS 493
 Db 365 GSDEERGEAEASSSTPATGVFLKSW-----VYQGEDTEEBEEDSDTGSADEREAE 417
 Qy 494 DSTGWSGCGVSGCGEGPLPETPDH-----SSGEEDDWEPSADAEENLKNWSPCHSE 547
 Db 418 TSASTPPASAFKAWVYRPGDTEEBEEDDVSDEKDDSEALGEAES----- 466
 Qy 548 DPYNLLNFKAPFQPSG-----KNWKGQDSKASSVTVAFSGHHTLLSCKAQLLESQDNC 603
 Db 467 DPH-----PSHPDQRAHFRGMYRPGKETEE-----EAEED-- 498
 Qy 604 PGCGLGEA-----LAGER-----YTHIKRK 623
 Db 499 ----WGEAEPFPRVAIYVPGKPPPPWAPPLRLQLRLLKRPETPTDPPETPLKAR 554
 Qy 624 KVTFLLEVTEYIIS-----GDEDRKGPWEFARDGCRFQKRIQETEAIGYCLAFHREK 678
 Db 555 KVRSEKVTYVHFLAVWAGPAQAARQGWELQARDRSFARRITQAQBELSPCLTPAAR 614
 Qy 679 MFNRLR 684
 Db 615 AWARLR 620

RESULT 14
 ADRI4089
 ID ADRI4089 standard; protein; 674 AA.
 XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 09:17:12 ; Search time 18.798 Seconds
(without alignments)
2771.838 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRARPGPRLGSWF.....MPNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	5.5	657	3	US-08-893-852A-3
2	204	5.5	657	3	US-08-821-818-3
3	204	5.5	657	4	US-09-052-753B-3
4	191.5	5.2	674	3	US-08-893-852A-1
5	182.5	4.9	578	4	US-09-052-753B-7
6	171.5	4.6	590	3	US-08-893-852A-4
7	171.5	4.6	590	3	US-08-821-818-2
8	171.5	4.6	590	4	US-09-052-753B-2
9	132.5	3.6	1065	4	US-09-949-016-11618
10	127.5	3.4	3913	4	US-09-949-016-10933
11	127	3.4	1503	3	US-08-976-255-14
12	125	3.4	2781	4	US-09-698-295-10
13	125	3.4	2907	4	US-09-698-295-1
14	124.5	3.4	1780	1	US-08-769-309A-5
15	124.5	3.4	1780	3	US-08-994-570-5
16	123.5	3.3	4377	4	US-09-949-016-6978
17	123	3.3	1142	2	US-08-993-118-7
18	123	3.3	1142	3	US-08-845-528C-7
19	123	3.3	1142	3	US-09-061-709-2
20	123	3.3	1142	4	US-09-066-281B-7
21	123	3.3	1142	4	US-09-899-651-2
22	123	3.3	1142	4	US-09-468-433C-7
23	123	3.3	1142	4	US-09-392-714-26
24	123	3.3	1142	4	US-09-270-437D-2
25	120.5	3.2	1133	4	US-09-252-991A-17945
26	120	3.2	1261	4	US-09-949-016-9651
27	119.5	3.2	961	4	US-09-538-092-1231

28	119	3.2	414	4	US-09-248-796A-19046	Sequence 19046, A
29	118	3.2	1781	4	US-09-961-403-13	Sequence 13, Appl
30	117.5	3.2	1712	4	US-09-949-016-9450	Sequence 9450, Ap
31	117	3.2	1234	2	US-08-317-310A-15	Sequence 15, Appl
32	117	3.2	1234	5	PCT-US95-13041-15	Sequence 15, Appl
33	117	3.2	1317	3	US-09-083-521-7	Sequence 7, Appl
34	116.5	3.1	912	5	PCT-US95-03747-2	Sequence 2, Appl
35	115	3.1	805	4	US-09-344-624-19	Sequence 19, Appl
36	115	3.1	805	4	US-09-949-016-6713	Sequence 6713, Ap
37	115	3.1	3340	4	US-09-252-991A-23568	Sequence 23568, A
38	114.5	3.1	654	1	US-08-083-590A-16	Sequence 16, Appl
39	114.5	3.1	654	2	US-08-346-128-35	Sequence 35, Appl
40	114.5	3.1	654	3	US-08-532-384-16	Sequence 16, Appl
41	114.5	3.1	1187	1	US-08-320-559-28	Sequence 28, Appl
42	114.5	3.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
43	114.5	3.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl
44	114.5	3.1	1210	1	US-08-320-559-26	Sequence 26, Appl
45	114.5	3.1	1210	3	US-08-545-860D-26	Sequence 26, Appl
46	114.5	3.1	1210	4	US-09-538-092-1179	Sequence 1179, Ap
47	114.5	3.1	1210	5	PCT-US94-04496-26	Sequence 26, Appl
48	114.5	3.1	2523	1	US-08-185-432-18	Sequence 18, Appl
49	114.5	3.1	2523	4	US-08-899-232-3	Sequence 3, Appl
50	114.5	3.1	2523	4	US-09-121-457-3	Sequence 3, Appl
51	114	3.1	605	3	US-09-394-645-2	Sequence 2, Appl
52	114	3.1	605	3	US-09-243-560B-2	Sequence 2, Appl
53	114	3.1	848	3	US-08-976-255-10	Sequence 10, Appl
54	114	3.1	1219	4	US-09-344-624-4	Sequence 4, Appl
55	113.5	3.1	445	3	US-08-845-258-38	Sequence 38, Appl
56	113.5	3.1	445	3	US-08-990-571-38	Sequence 38, Appl
57	113.5	3.1	445	3	US-08-723-142A-38	Sequence 38, Appl
58	113.5	3.1	445	4	US-09-528-784A-38	Sequence 38, Appl
59	113.5	3.1	445	4	US-09-569-098A-38	Sequence 38, Appl
60	113.5	3.1	605	4	US-09-949-016-11347	Sequence 11347, A
61	113.5	3.1	666	4	US-09-528-784A-85	Sequence 85, Appl
62	113.5	3.1	666	4	US-09-569-098A-85	Sequence 85, Appl
63	113.5	3.1	1132	4	US-09-528-784A-87	Sequence 87, Appl
64	113.5	3.1	1132	4	US-09-569-098A-87	Sequence 87, Appl
65	112.5	3.0	1076	1	US-09-949-016-7421	Sequence 7421, Ap
66	112.5	3.0	1805	1	US-07-853-913-2	Sequence 2, Appl
67	112	3.0	576	2	US-09-367-206-22	Sequence 22, Appl
68	112	3.0	1183	3	US-08-447-031A-2	Sequence 2, Appl
69	112	3.0	2353	3	US-08-984-709A-50	Sequence 50, Appl
70	111.5	3.0	377	4	US-09-248-796A-20227	Sequence 20227, A
71	111.5	3.0	619	3	US-09-066-046-2	Sequence 2, Appl
72	111.5	3.0	773	4	US-09-270-767-33141	Sequence 33141, A
73	111.5	3.0	773	4	US-09-270-767-48358	Sequence 48358, A
74	111.5	3.0	1969	4	US-09-418-710-72	Sequence 72, Appl
75	111.5	3.0	1969	4	US-09-839-479-71	Sequence 71, Appl
76	111	3.0	576	3	US-09-367-206-1	Sequence 1, Appl
77	111	3.0	576	3	US-09-367-206-21	Sequence 21, Appl
78	111	3.0	576	3	US-09-367-206-23	Sequence 23, Appl
79	111	3.0	581	4	US-09-244-805-27	Sequence 27, Appl
80	111	3.0	1155	1	US-08-094-948A-29	Sequence 29, Appl
81	111	3.0	1155	5	PCT-US96-03319-29	Sequence 29, Appl
82	110.5	3.0	2870	4	US-09-479-467A-15	Sequence 15, Appl
83	110.5	3.0	3178	4	US-09-479-467A-4	Sequence 4, Appl
84	110	3.0	484	3	US-09-066-046-8	Sequence 8, Appl
85	110	3.0	489	1	US-07-903-103-4	Sequence 4, Appl
86	110	3.0	489	1	US-08-044-619A-4	Sequence 4, Appl
87	110	3.0	489	1	US-08-283-911-4	Sequence 4, Appl
88	110	3.0	489	1	US-08-245-500A-5	Sequence 5, Appl
89	110	3.0	489	1	US-08-390-546-5	Sequence 5, Appl
90	110	3.0	489	1	US-08-390-479A-5	Sequence 5, Appl
91	110	3.0	489	1	US-08-857-393-5	Sequence 5, Appl
92	110	3.0	489	1	US-08-350-516C-5	Sequence 5, Appl
93	110	3.0	489	1	US-08-390-517A-5	Sequence 5, Appl
94	110	3.0	489	2	US-08-801-718-5	Sequence 5, Appl
95	110	3.0	489	2	US-08-801-718-5	Sequence 5, Appl
96	110	3.0	489	3	US-09-170-159A-5	Sequence 5, Appl
97	110	3.0	489	4	US-09-480-718-46	Sequence 46, Appl
98	110	3.0	544	3	US-09-087-134-14	Sequence 14, Appl
99	110	3.0	757	4	US-09-949-016-6963	Sequence 6963, Ap
100	110	3.0	758	4	US-09-949-016-8087	Sequence 8087, Ap

ALIGNMENTS

RESULT 1

US-08-893-852A-3
 ; Sequence 3, Application US/08893852A
 ; Patent No. 6080558
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: Fast-Seq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/893,852A
 ; FILING DATE: Filed Herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0341 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 657 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 53041
 ; US-08-893-852A-3

Query Match 5.5%; Score 204; DB 3; Length 657;

Best Local Similarity 21.5%; Pred. No. 9.4e-10;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

QY 60 LLSQALLPSLFOKLLWSGLIPTRWLDFAASYSALRASGRRESDAPTQKSL 119
 DB 21 LLSPLMGLLR-----NWSRLRGPEVPEAWL-----AKTVTGADQ 55
 QY 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLRLVLRLEVKLKAQERAL---DSAAPTFL 171
 DB 56 IEMALLPTPTVSGNL-LPHGETESGSP-----EQSQAQRCLCLVEAESPP----- 102
 QY 172 EQQLWGVLLPS-SIQAGLVSHRELD-SSSSGPLSVQSLG-----NFKVSVY---LLN 219
 DB 103 --ETWGLSNVDEYNAPGQDRLREKEMERTAGKATLPAGLQAGADKRLGEVVAEEGVAE 160
 QY 220 PSYLDLPQLGLRCQSSAGGQFVGFRLLTPESCVLSYSDGCHPOPLRAEMSATAWRCPP 279
 DB 161 PAY-----PTSQLEGGPAEN-----EEDGETVTKYQASAASIAPGYKPS 199
 QY 280 LSTGLPEIHHRRMRWLVPFQNOGQDLPTLDQONGYHSLSEHNLLRMDPQHCTDNPQ 339

DB 200 TPVPFLGEAEHQATE-----EKGTEKADPSNPSGSGSHRAWY-YGREKPKQGEAKVE 254
 QY 340 AVSPAADRP-----EPTEKKPELVIQEYSQSPQSSLFCELPVEKECEDHTNATDLSDRG 395
 DB 255 AHRAGQGHPCRNABAEAGGPETTF--VCTGNAFKAWVYRPGEDTEEDNSDSASEDT 312
 QY 396 ESLPVSSTRPVCNKLDIYILGGAPSDLEASDSSESD-----W---GE 435
 DB 313 AQTGATHTSAFLKAWVYRPGEDTEEDSDSASEDTAQTGATHTSAFLKAWVYRPG 372
 QY 436 EPEDDGFSDGSLSESDVEQDSEGLHLWNSFHSVDPYKQNFATTIATAARIAPRPSDS 495
 DB 373 DTEENSDDLDS--AEEDTAQTGATHT--SAFLKAWVYRPGEDTEEDNSDLDSASEDTAQT 429
 QY 496 GTSWSGSGVSGCOEGPLPETPHSSGCEEDWEDSAEENKLWNSFCHSEDPYNLLNP 555
 DB 430 GATHTTSPFLKAWVYRPGEDTDEEDSEENAVPGDSETADSSQSPC----- 478
 QY 556 KAPFQPS-----GKNWKGQDSKASSEVTVAFSGHH-----TLLSCKAQLLES-- 598
 DB 479 ---LQFORCLPGEKTKGRGEPPLFQVAFYLPGEKPSWAAPKLPLRLQRLFLKAPT 535
 QY 599 --QEDNCPGCGLGEALAGERYTHIKKXVTFLEEVTEYYS-----GDEDRKGPWESFAR 651
 DB 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPEWQFAR 580
 QY 652 DGCRRFQRIQETETVAIGYCLAFEHREKMFNRLR 684
 DB 581 DRSRFAIRIAQAEEKLGLPYLTPOSRARAWALR 613

RESULT 2

US-08-821-818-3
 ; Sequence 3, Application US/08821818
 ; Patent No. 6146877
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Paul B.
 ; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
 ; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/821,818
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Chan, Albert Wai-Kit
 ; REGISTRATION NUMBER: 36,479
 ; REFERENCE/DOCKET NUMBER: 51523
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 657 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-821-818-3

Query Match 5.5%; Score 204; DB 3; Length 657;
Best Local Similarity 21.5%; Pred. No. 9.4e-10;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;
Qy 60 LLSQALLPSLFQKLLWSQSGGLPTRWLDPFAASYSALRASRGRESDAPTVQKSL 119
Db 21 LLSPLMGLLR-----AWSRLRGPEVPEAWL-----AKTVTGADQ 55
Qy 120 YTAAGLFAKTRVSTLALRG-----GTPVAVLVLRLEVKLKAQERAL-----DSAAPTFL 171
Db 56 IEAALTLPTPVSGNL-LPHGETESGSP-----EQSQAQRLCLVEAESPP----- 102
Qy 172 EQQLWGVLLPS-SLOAGLVSHRELD-SSSGPLSVQSLG-----NFKVVS-LLN 219
Db 103 --ETWGLSNVDEYNAKPGQDRLREKEMERTAGKATLQAGLOGADKRLGEVVARREGVAE 160
Qy 220 PSYLDYLPQLGLRCQSSAGGQFVGFRTLTPESCYLSSEDCGHPQPLRAEMSATAWRCPP 279
Db 161 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASTIAPGYKPS 199
Qy 280 LSTEGLEPIHRRMRMLVFLQPNQODLPTLDQDNGYHSLSEEHNLRLMDPQHCTDNPAQ 339
Db 200 TPVPFLGEAEHQATE-----EKGTEKADPNPSPSGSHSRAWEY-YREKPKQGEAKVE 254
Qy 340 AVSPAADRP-----EPTKPKPELVIOEVSQSPQSGSLFCELPVEKECEDHTNATDLSDRG 395
Db 255 AHRAGQGHPCRNAAEAGGPETTF--VCTGNAPFLKAWVYRPGEDTEEDNSDSAEEDT 312
Qy 396 ESLPSTRPVCSNKLIDYILGAPSDLEASDSEED-----W-----GE 435
Db 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSAEEDTAQTGATPHTSAFLKAWVYRPG 372
Qy 436 EPDDGDSGLSESVDQSEGLHLWNSFHSVDYKPNQFTATITAAARIAPRDPDS 495
Db 373 DTEENSGLDS--AEEDTAQTGATPHT-SAPFLKAWVYRPGEDTEEDNSDLSAEEDTAQT 429
Qy 496 GTSWSGCGVSGCGEPLPETHDSSGEEDDWEPSADEAENLKLWNSFCHSEDPYNLNF 555
Db 430 GATPHTSPFLKAWVYRPGEDTDDTEEDSENVAPODSETADSSQSPC----- 478
Qy 556 KAPFQPS-----GKNWKGROQDSKASSEVTVAFSGH-----TLLSCKAQLLS-- 598
Db 479 --LQPCRLPGERTKGRGEEPLFQVAFYLPGEKPSWPAAPKLPRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGLGEALAGERYTHIKRKVTFLFEVTEYIS-----GDEDRKGPWEERFAR 651
Db 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPEWQFAR 580
652 DGCRCFKRIQETVAIGYCLAFEHREKMFNRLR 684
581 DRSRFAIRIAQAEKLGPLYLTPDSRARAWARL 613

RESULT 3

US-09-052-753B-3

; Sequence 3, Application US/09052753B

; Patent No. 6472520

; GENERAL INFORMATION:

; APPLICANT: Paul B. Fisher

; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses

; FILE REFERENCE: A34608-B

; CURRENT APPLICATION NUMBER: US/09/052,753B

; PRIOR FILING DATE: 1998-03-31

; PRIOR APPLICATION NUMBER: PCT/US98/05793

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 08/812,818

; PRIOR FILING DATE: 1997-03-21

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 657

; TYPE: PRT

; ORGANISM: rat
US-09-052-753B-3

Query Match 5.5%; Score 204; DB 4; Length 657;

Best Local Similarity 21.5%; Pred. No. 9.4e-10;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

Qy 60 LLSQALLPSLFQKLLWSQSGGLPTRWLDPFAASYSALRASRGRESDAPTVQKSL 119
Db 21 LLSPLMGLLR-----AWSRLRGPEVPEAWL-----AKTVTGADQ 55
Qy 120 YTAAGLFAKTRVSTLALRG-----GTPVAVLVLRLEVKLKAQERAL-----DSAAPTFL 171
Db 56 IEAALTLPTPVSGNL-LPHGETESGSP-----EQSQAQRLCLVEAESPP----- 102
Qy 172 EQQLWGVLLPS-SLOAGLVSHRELD-SSSGPLSVQSLG-----NFKVVS-LLN 219
Db 103 --ETWGLSNVDEYNAKPGQDRLREKEMERTAGKATLQAGLOGADKRLGEVVARREGVAE 160
Qy 220 PSYLDYLPQLGLRCQSSAGGQFVGFRTLTPESCYLSSEDCGHPQPLRAEMSATAWRCPP 279
Db 161 PAY-----PTSQLEGGPAEN-----BEDGETVKTYQASAASTIAPGYKPS 199
Qy 280 LSTEGLEPIHRRMRMLVFLQPNQODLPTLDQDNGYHSLSEEHNLRLMDPQHCTDNPAQ 339
Db 200 TPVPFLGEAEHQATE-----EKGTEKADPNPSPSGSHSRAWEY-YREKPKQGEAKVE 254
Qy 340 AVSPAADRP-----EPTKPKPELVIOEVSQSPQSGSLFCELPVEKECEDHTNATDLSDRG 395
Db 255 AHRAGQGHPCRNAAEAGGPETTF--VCTGNAPFLKAWVYRPGEDTEEDNSDSAEEDT 312
Qy 396 ESLPSTRPVCSNKLIDYILGAPSDLEASDSEED-----W-----GE 435
Db 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSAEEDTAQTGATPHTSAFLKAWVYRPG 372
Qy 436 EPDDGDSGLSESVDQSEGLHLWNSFHSVDYKPNQFTATITAAARIAPRDPDS 495
Db 373 DTEENSGLDS--AEEDTAQTGATPHT-SAPFLKAWVYRPGEDTEEDNSDLSAEEDTAQT 429
Qy 496 GTSWSGCGVSGCGEPLPETHDSSGEEDDWEPSADEAENLKLWNSFCHSEDPYNLNF 555
Db 430 GATPHTSPFLKAWVYRPGEDTDDTEEDSENVAPODSETADSSQSPC----- 478
Qy 556 KAPFQPS-----GKNWKGROQDSKASSEVTVAFSGH-----TLLSCKAQLLS-- 598
Db 479 --LQPCRLPGERTKGRGEEPLFQVAFYLPGEKPSWPAAPKLPRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGLGEALAGERYTHIKRKVTFLFEVTEYIS-----GDEDRKGPWEERFAR 651
Db 536 RDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPEWQFAR 580
652 DGCRCFKRIQETVAIGYCLAFEHREKMFNRLR 684
581 DRSRFAIRIAQAEKLGPLYLTPDSRARAWARL 613

RESULT 4

US-08-893-852A-1

; Sequence 1, Application US/08893852A

; Patent No. 6080558

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; APPLICANT: Shah, Purvi

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

Qy 597 ESQEDNCPGCGLGGALAGERYTHIKRKKVTFLEEVTEYYIS-----GDEDRKGWEEFAR 651
 Db 448 -----PARNQGPFI-----LKGKRVHFSEKVTVHFLAVWAGPAQAARRGPEWQFAR 494
 Qy 652 DGCRRFQRIQETVAIGYCLAFEHREKMFRLR 684
 Db 495 DRSRFRARRIAQAEEQLGPLYLTPAFRARAATRLR 527

RESULT 6
 US-08-893-852A-4
 ; Sequence 4, Application US/08893852A
 ; Patent No. 6080558
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/893,852A
 ; FILING DATE: Filed Herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PP-0341 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 590 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 452490
 ; US-08-893-852A-4

Query Match 4.6%; Score 171.5; DB 3; Length 590;
 Best Local Similarity 18.1%; Pred. No. 8.1e-07;
 Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSEDCGHPQLRAEMSAATARRCPPL-----STEGLEPIHRR 292
 Db 103 PETGLSDDDKQGDGPREQGRAHTAGLPIILSPGLQSAKSLGVVAGEGVTELAYPT 162
 Qy 293 MRWLVLQPNQGDLPILDQNG-----YHLEEEH 323
 Db 163 SHW-----EGCPSEEDGETVKKAFRASADSPGHKSTSVYCPGEAHOATEBK 213
 Qy 324 NLLRWDP-----QHCTDNPAQVSPADRP-----TEKKPELV 358
 Db 214 TENKADPPSPSGSHSRAMEYCSKQEGEA-----DPEPHRAGKYQLCQNAEAESEAK 267
 Qy 359 IQEVSQSPQSSSL--FCLELPVEKECEDHTNATDLSDRGESLPVSTRPVCNKLIDYILG 416

Db 268 VSSLSVSSGNFLKAWVYRPGEDTDDDDSDWGSAAEEGKALSSPTSP-----HDPLKA 322
 Qy 417 GAPSDLEASDSSESDWGEPEDDGDFDSGLSSESVDVEODSEGLHLWNSFHSVDPYKPN 476
 Db 323 WYVRPGEDTDDDDSDWGSAAE-----EGKALSSPTSPEHDFLKAW-----VYRPG 369
 Qy 477 FTATIQTAAARIAPRDPDSGTSWGSVCGVSCQEGPLPET-----516
 Db 370 DT-----EDDQSDWGSAAEKDGLAQTFATPHTSAFLKTVWCPCGDETDGDC 416
 Qy 517 -----PDHSGEEDDWEPS-----ADEAENLKLWNSFCHSEDPYNLLNFKAPFPQSGK 564
 Db 417 EVVVPEDSEAADPKSPSHEAQCLPGEQTEGL-----VEAHSLSFQVAFYLPGEKPA 470
 Qy 565 NWKGRQDSKASSEVTVAFGSHHTLLSCKAQLLESQEDNCPGCGLGGALAGERYTHIKRKK 624
 Db 471 PW-----TAPKLPRLQRLTLRTPTQ---DQDPETP-----LRARK 505
 Qy 625 VTFLEEVTEYYIS-----GDEDRKGWEEFARDCGRFQRIQETVAIGYCLAFEHREK 679
 Db 506 VHFSENVTVHFLAVWAGPAQAARRGPEWQELARDRSRFRARRIAQAEEKLGPLYLTPAFRARA 565
 Qy 680 FNRRLRIESKDLLL 692
 Db 566 WARLGNPSLPLAL 578

RESULT 7

US-08-821-818-2
 ; Sequence 2, Application US/08821818
 ; Patent No. 6146877
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Paul B.
 ; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
 ; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/821,818
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Chan, Albert Wai-Kit
 ; REGISTRATION NUMBER: 36,479
 ; REFERENCE/DOCKET NUMBER: 51523
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 590 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-821-818-2

Query Match 4.6%; Score 171.5; DB 3; Length 590;
 Best Local Similarity 18.1%; Pred. No. 8.1e-07;
 Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

[illegible]

```

RESULT 10
US-09-949-016-10933
; Sequence 10933, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10933
; LENGTH: 3913
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10933

```

Query Match 3.4%; Score 127.5; DB 4; Length 3913;
Best Local Similarity 19.1%; Pred. No. 0.2;
Matches 159; Conservative 113; Mismatches 346; Indels 211; Gaps 33;

QY	5	THRARKRPGRLGSRFLPLRLRSHACSEFP	PPSSRQNPNGALPERRTRYTKLLSQL	64
Db	1			
Db	1155	TFSSRTSPVTTAGS----	LLERSITMTTPASPKNINMYSSSLP-----FKSIITSA	1203
QY	65	LALLPSLFQ-----	KLILWSQLS-----	92
Db	1			
Db	1204	APLISGLKVVSPVKSADVVISAKITHASSLS	SPVKQMPGHAENVLNGSISPLKYP	1263
QY	93	FAASYGALRASRGRESDAPTVQKSLSYTAAGL	FAKTRVVVSTLALARGGTPVAVLVRLE	152
Db	1			
Db	1264	SSTLINGCKATATLQEKISSATNSVS	SVSAATDTVEKVPSTTT-----AMPPGPL	1314
QY	153	VKLKAQERALDSPAATFELLEQQLGWGVELLP	SSLSQAGLVSHRELDSSSGSGLSPVQSLGNFK	212
Db	1			
Db	1315	-----RSVSAAPSAFOSLRTPSASALYTS	LSGISA-----TSSVTSIIITVPVYS	1362
QY	213	VWYLLNPSYLDYLPQGLRCQSSAGGQGVG	FRTILTPBSCYLSEDCGCHPQPLRAEMSA-	271
Db	1			
Db	1363	VWVLPPEPA-LKKLPDSNFTKSAA-	ALLSPIKTLTET-----HPQHPSTRSSP	1411
QY	272	-----TAWRRCPPLS-----	TEGLPEIHHRRMRMLVFLQNGQDLP-----	308
Db	1			
Db	1412	VKSGLFAPSALKLSTPSSLSQELIKDVAEM	KEDLMRTAILQTDVPEKFPQPELPK	1471
QY	309	--TLDQNGVHSLSE-ENHLR----	MDPOHCTDNPAQAVSPAADR-PEPTEKKP	360
Db	1			
Db	1472	EGRIDDEEPKIVKVKBDLVKVSILKDKDV	CNDKNGSPKSPKSGHSPEDDOWIEPSE	1531
QY	361	EV-----SOSPQ-----	-----GSSLFCBL-----P	376
Db	1			
Db	1532	EIREARQAAAASQSPSLPERVOVKAKAASE	KDYNLTKVIDYLTNDIGSSSLTNLYK	1591
QY	377	VEKECEDHTN-----ATDLSDRGESLP-	-----VSTRPVCSNKLIDYILGGAPS	425
Db	1			
Db	1592	AKKDGBERQKVLKPAITALQEHKLKMP	PASMRSTSEKELC--KMAДСFF-GTDTIL	1648
QY	426	SDSESDWGEEP-EDDGFSDGSLSDSDVQD	SEGLHLWNPSHSDVPYKPKQNFATTIQT	484
Db	1			
Db	1649	DDFSQHDQDKSPJSDSGFETRSKTPS-	APQASBSTGPKPLFHEV-PIPPVITR	1706
QY	485	ARIAPRDPDSGTSWSGSCGVGSCQGLP	PETPDHSSGEEDDWEPSADEAENLKL	544
Db	1			
Db	1707	HVIRSYDPSAGDVPQIQ-----	PEEPVSPKSPPTFMELEPKPTTSSIKE	1759
QY	545	HS-EDPN-----	LNFKAPFPQSGKNWGR-QDSKASSEVT	583
Db	1			
Db	1760	SSEDDHNRLVLSKGMVKEETHITTTTR	VYHSP--PGGSGASERIEBTMSVHDIM	1817
QY	584	GHHTLLSCKAQLLESQEDNCPGCGLGA	LAGERYTHIKRKVFLREBTVYITSGDE	643
Db	1			
Db	1818	SGRDPSELAGLFEHKASVSPDVHK	GAETSAGAEKDNQMKPLERIEVHI----	1872
QY	644	GPWEEFARDGCRFQKRIQETEVAIGY	CLAFEHREKMPNRLRIESKOLL	691
Db	1			
Db	1873	GNQAEPTVEIIRETKHPEKEMVY-	-----YQKDLSRGDINLKDPL	1912
RESULT 11				
US-08-976-255-14				
; Sequence 14, Application US/08976255				
; Patent No. 6136581				
; GENERAL INFORMATION:				
; APPLICANT: Jono, Keith E.				
; APPLICANT: Plowman, Gregory				
; TITLE OF INVENTION: KINASE GENES AND USES				
; NUMBER OF SEQUENCES: 53				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Lyon & Lyon				
; STREET: 633 West Fifth Street				
; STREET: Suite 4700				
; CITY: Los Angeles				
; STATE: California				

```

RESULT 11
US-08-976-255-14
Sequence 14, Application US/08976255
Patent No. 6136581
GENERAL INFORMATION:
APPLICANT: Jono, Keith E.
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: KINASE GENES
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSER: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California

```


Qy	99	ALRASRGRESD	---	APT	V	K	S	L	S	Y	T	A	G	L	F	A	K	T	R	V	V	S	T	L	A	L	R	G	T	P	V	A	V	L	R	E	V	K	L	155
Db	861	AQQAQGAEPQ	E	K	A	A	T	E	V	S	K	E	S	Q	V	H	M	A	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	905
Qy	156	KAQERALD	---	S	A	A	P	T	F	L	E	Q	L	W	C	V	E	L	L	P	S	S	L	O	A	G	L	V	S	H	R	E	L	D	S	S	S	Q	153	
Db	906	--PERSPSWIS	A	S	A	S	T	E	P	E	Q	---	V	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	940		
Qy	214	VSYLNP	S	I	D	Y	I	P	O	L	G	R	C	O	S	S	A	G	G	O	F	V	G	F	R	T	L	T	P	E	S	C	V	L	S	B	D	G	268	
Db	941	IABEPP	T	V	T	E	P	L	P	E	---	N	R	E	A	R	G	D	T	V	S	E	A	L	T	P	E	A	V	A	E	T	A	---	---	---	---	992		
Qy	269	---	---	M	S	A	T	A	R	R	C	P	P	L	S	T	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	309			
Db	993	SAABET	T	E	M	V	S	A	V	S	Q	L	T	D	S	P	D	T	T	E	A	T	P	V	O	E	V	E	G	V	D	I	E	B	E	Q	---	1046		
Qy	310	LDQNG	V	H	S	L	E	E	E	H	N	L	M	P	D	P	O	H	C	T	D	N	P	A	O	V	S	P	---	---	---	---	---	---	---	---	359			
Db	1047	---	---	E	K	V	K	E	S	Q	---	---	P	O	T	G	P	E	D	V	L	P	V	Q	R	A	E	A	E	R	P	E	Q	A	S	L	---	1093		
Qy	360	---	---	Q	E	V	S	O	S	P	---	---	C	G	S	L	F	C	E	L	P	E	K	E	C	E	D	H	T	N	A	T	L	D	---	---	---	407		

Search completed: September 16, 2005, 10:47:19
Job time : 20.798 secs

```

; REFERENCE/DOC#1 NUMBER: 27868/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-994-570-5

Query Match 3.4%; Score 124.5; DB 3; Length 1780;
Best Local Similarity 18.9%; Pred. No. 0.11;
Matches 153; Conservative 96; Mismatches 237; Indels 319; Gaps 38;

QY 99 ALRASRGRESD---APTQKSLSYTAAGLFAKTRVVSTLALARGGTPVAVLVLRLEVKL 155
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 861 AQAQKGAEQEQAATEVSKELSESQVHMA-----AAVADGTRAATII----- 905

QY 156 KAQERALD--SAAPTFLLEQLWGVLLPSSLQAGLVSHRELDSSSSCPLSVQSLGNPKV 213
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 906 --BERSPSWISASVTEPLEQ---VE-----AEAALLTEEVLER-----EV 940

QY 214 VSYLLNPSYDLYLPQLGLRCQSSAGGQGVFGFRTLTPESCYSVLSDGCHPOPLRAE----- 268
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 941 IAEETPTVTEPLE-----NREAGDTVSEALTPENVTAETA---GPLSSEGTGA 992

QY 269 -----NSATAWRRCPLPSTE-----GLPEIHRRMRWLVTLPQNGQDLPT 309
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 993 SAAAEETPMVSAVSQTDSPTTTEATPVQVEGGVPDIEQEERTQEVQLQAVA----- 1046

QY 310 LDQDNGVHSLSEEHNLRMDFQHCTDNPQAQVSPA---ADRPF-----TEKKPELVI 359
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1047 -----EKVKESQL-----PQTGGFVDVLQPVQRAEAERPEQAEASGLKKETDVL 1093

QY 360 ----QEVSQSP--CGSSLFCFELPVEKECEEDHTNATLSD-----RGESLP-VSTRPVCS 407

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:46:08 ; Search time 237.943 Seconds
(without alignments)
1187.844 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRKRPGLGSWF.....MFNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 404927589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 100 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US10G_PUBCOMB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 21: /cgn2_6/ptodata/1/pubpaa/US11A_NEW_PUB.pep.*
 - 22: /cgn2_6/ptodata/1/pubpaa/US11B_NEW_PUB.pep.*
 - 23: /cgn2_6/ptodata/1/pubpaa/US11C_NEW_PUB.pep.*
 - 24: /cgn2_6/ptodata/1/pubpaa/US11D_NEW_PUB.pep.*
 - 25: /cgn2_6/ptodata/1/pubpaa/US11E_NEW_PUB.pep.*
 - 26: /cgn2_6/ptodata/1/pubpaa/US11F_NEW_PUB.pep.*
 - 27: /cgn2_6/ptodata/1/pubpaa/US11G_NEW_PUB.pep.*
 - 28: /cgn2_6/ptodata/1/pubpaa/US11H_NEW_PUB.pep.*
 - 29: /cgn2_6/ptodata/1/pubpaa/US11I_NEW_PUB.pep.*
 - 30: /cgn2_6/ptodata/1/pubpaa/US11J_NEW_PUB.pep.*
 - 31: /cgn2_6/ptodata/1/pubpaa/US11K_NEW_PUB.pep.*
 - 32: /cgn2_6/ptodata/1/pubpaa/US11L_NEW_PUB.pep.*
 - 33: /cgn2_6/ptodata/1/pubpaa/US11M_NEW_PUB.pep.*
 - 34: /cgn2_6/ptodata/1/pubpaa/US11N_NEW_PUB.pep.*
 - 35: /cgn2_6/ptodata/1/pubpaa/US11O_NEW_PUB.pep.*
 - 36: /cgn2_6/ptodata/1/pubpaa/US11P_NEW_PUB.pep.*
 - 37: /cgn2_6/ptodata/1/pubpaa/US11Q_NEW_PUB.pep.*
 - 38: /cgn2_6/ptodata/1/pubpaa/US11R_NEW_PUB.pep.*
 - 39: /cgn2_6/ptodata/1/pubpaa/US11S_NEW_PUB.pep.*
 - 40: /cgn2_6/ptodata/1/pubpaa/US11T_NEW_PUB.pep.*
 - 41: /cgn2_6/ptodata/1/pubpaa/US11U_NEW_PUB.pep.*
 - 42: /cgn2_6/ptodata/1/pubpaa/US11V_NEW_PUB.pep.*
 - 43: /cgn2_6/ptodata/1/pubpaa/US11W_NEW_PUB.pep.*
 - 44: /cgn2_6/ptodata/1/pubpaa/US11X_NEW_PUB.pep.*
 - 45: /cgn2_6/ptodata/1/pubpaa/US11Y_NEW_PUB.pep.*
 - 46: /cgn2_6/ptodata/1/pubpaa/US11Z_NEW_PUB.pep.*
 - 47: /cgn2_6/ptodata/1/pubpaa/US12_NEW_PUB.pep.*
 - 48: /cgn2_6/ptodata/1/pubpaa/US12A_NEW_PUB.pep.*
 - 49: /cgn2_6/ptodata/1/pubpaa/US12B_NEW_PUB.pep.*
 - 50: /cgn2_6/ptodata/1/pubpaa/US12C_NEW_PUB.pep.*
 - 51: /cgn2_6/ptodata/1/pubpaa/US12D_NEW_PUB.pep.*
 - 52: /cgn2_6/ptodata/1/pubpaa/US12E_NEW_PUB.pep.*
 - 53: /cgn2_6/ptodata/1/pubpaa/US12F_NEW_PUB.pep.*
 - 54: /cgn2_6/ptodata/1/pubpaa/US12G_NEW_PUB.pep.*
 - 55: /cgn2_6/ptodata/1/pubpaa/US12H_NEW_PUB.pep.*
 - 56: /cgn2_6/ptodata/1/pubpaa/US12I_NEW_PUB.pep.*
 - 57: /cgn2_6/ptodata/1/pubpaa/US12J_NEW_PUB.pep.*
 - 58: /cgn2_6/ptodata/1/pubpaa/US12K_NEW_PUB.pep.*
 - 59: /cgn2_6/ptodata/1/pubpaa/US12L_NEW_PUB.pep.*
 - 60: /cgn2_6/ptodata/1/pubpaa/US12M_NEW_PUB.pep.*
 - 61: /cgn2_6/ptodata/1/pubpaa/US12N_NEW_PUB.pep.*
 - 62: /cgn2_6/ptodata/1/pubpaa/US12O_NEW_PUB.pep.*
 - 63: /cgn2_6/ptodata/1/pubpaa/US12P_NEW_PUB.pep.*
 - 64: /cgn2_6/ptodata/1/pubpaa/US12Q_NEW_PUB.pep.*
 - 65: /cgn2_6/ptodata/1/pubpaa/US12R_NEW_PUB.pep.*
 - 66: /cgn2_6/ptodata/1/pubpaa/US12S_NEW_PUB.pep.*
 - 67: /cgn2_6/ptodata/1/pubpaa/US12T_NEW_PUB.pep.*
 - 68: /cgn2_6/ptodata/1/pubpaa/US12U_NEW_PUB.pep.*
 - 69: /cgn2_6/ptodata/1/pubpaa/US12V_NEW_PUB.pep.*
 - 70: /cgn2_6/ptodata/1/pubpaa/US12W_NEW_PUB.pep.*
 - 71: /cgn2_6/ptodata/1/pubpaa/US12X_NEW_PUB.pep.*
 - 72: /cgn2_6/ptodata/1/pubpaa/US12Y_NEW_PUB.pep.*
 - 73: /cgn2_6/ptodata/1/pubpaa/US12Z_NEW_PUB.pep.*
 - 74: /cgn2_6/ptodata/1/pubpaa/US13_NEW_PUB.pep.*
 - 75: /cgn2_6/ptodata/1/pubpaa/US13A_NEW_PUB.pep.*
 - 76: /cgn2_6/ptodata/1/pubpaa/US13B_NEW_PUB.pep.*
 - 77: /cgn2_6/ptodata/1/pubpaa/US13C_NEW_PUB.pep.*
 - 78: /cgn2_6/ptodata/1/pubpaa/US13D_NEW_PUB.pep.*
 - 79: /cgn2_6/ptodata/1/pubpaa/US13E_NEW_PUB.pep.*
 - 80: /cgn2_6/ptodata/1/pubpaa/US13F_NEW_PUB.pep.*
 - 81: /cgn2_6/ptodata/1/pubpaa/US13G_NEW_PUB.pep.*
 - 82: /cgn2_6/ptodata/1/pubpaa/US13H_NEW_PUB.pep.*
 - 83: /cgn2_6/ptodata/1/pubpaa/US13I_NEW_PUB.pep.*
 - 84: /cgn2_6/ptodata/1/pubpaa/US13J_NEW_PUB.pep.*
 - 85: /cgn2_6/ptodata/1/pubpaa/US13K_NEW_PUB.pep.*
 - 86: /cgn2_6/ptodata/1/pubpaa/US13L_NEW_PUB.pep.*
 - 87: /cgn2_6/ptodata/1/pubpaa/US13M_NEW_PUB.pep.*
 - 88: /cgn2_6/ptodata/1/pubpaa/US13N_NEW_PUB.pep.*
 - 89: /cgn2_6/ptodata/1/pubpaa/US13O_NEW_PUB.pep.*
 - 90: /cgn2_6/ptodata/1/pubpaa/US13P_NEW_PUB.pep.*
 - 91: /cgn2_6/ptodata/1/pubpaa/US13Q_NEW_PUB.pep.*
 - 92: /cgn2_6/ptodata/1/pubpaa/US13R_NEW_PUB.pep.*
 - 93: /cgn2_6/ptodata/1/pubpaa/US13S_NEW_PUB.pep.*
 - 94: /cgn2_6/ptodata/1/pubpaa/US13T_NEW_PUB.pep.*
 - 95: /cgn2_6/ptodata/1/pubpaa/US13U_NEW_PUB.pep.*
 - 96: /cgn2_6/ptodata/1/pubpaa/US13V_NEW_PUB.pep.*
 - 97: /cgn2_6/ptodata/1/pubpaa/US13W_NEW_PUB.pep.*
 - 98: /cgn2_6/ptodata/1/pubpaa/US13X_NEW_PUB.pep.*
 - 99: /cgn2_6/ptodata/1/pubpaa/US13Y_NEW_PUB.pep.*
 - 100: /cgn2_6/ptodata/1/pubpaa/US13Z_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3710	100.0	698	16	US-10-650-482-4
2	2223	59.9	713	16	US-10-650-482-2
3	2157.5	58.2	707	9	US-09-925-299-941
4	2157.5	58.2	707	10	US-09-925-299-941
5	547	14.7	153	15	US-10-296-115-963
6	204	5.5	657	9	US-09-052-753-3
7	204	5.5	657	15	US-10-277-603-3
8	186.5	5.0	674	16	US-10-755-889-90
9	182.5	4.9	578	9	US-09-052-753-7
10	182.5	4.9	578	15	US-10-277-603-7
11	171.5	4.6	590	9	US-09-052-753-2
					Sequence 4, Appli
					Sequence 2, Appli
					Sequence 941, App
					Sequence 941, App
					Sequence 963, App
					Sequence 3, Appli
					Sequence 30, Appli
					Sequence 90, Appli
					Sequence 7, Appli
					Sequence 7, Appli
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 38378, A
					Sequence 207028, A
					Sequence 66899, A
					Sequence 152, App
					Sequence 1354, App
					Sequence 36, Appli
					Sequence 98, Appli
					Sequence 130, App
					Sequence 15, App
					Sequence 6, Appli
					Sequence 10, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 1357, App
					Sequence 1355, App
					Sequence 1356, App
					Sequence 1024, App
					Sequence 201156, A
					Sequence 8429, App
					Sequence 2, Appli
					Sequence 684, App
					Sequence 316809, A
					Sequence 47, Appli
					Sequence 182739, A
					Sequence 38319, A
					Sequence 262, App
					Sequence 45, Appli
					Sequence 501, App
					Sequence 2964, App
					Sequence 4, Appli
					Sequence 2, Appli
					Sequence 490, App
					Sequence 8, Appli
					Sequence 14, Appli
					Sequence 487, App
					Sequence 141397, A
					Sequence 7, Appli
					Sequence 9, Appli
					Sequence 67803, A
					Sequence 122, App
					Sequence 1472, App
					Sequence 10, Appli
					Sequence 1, Appli
					Sequence 165014, A
					Sequence 41872, A
					Sequence 219612, A
					Sequence 340752, A
					Sequence 51377, A
					Sequence 4, Appli
					Sequence 3131, A
					Sequence 4917, App
					Sequence 37734, A
					Sequence 52303, A
					Sequence 247745, A
					Sequence 3123, App
					Sequence 2, Appli
					Sequence 79, Appli
					Sequence 73, Appli
					Sequence 54525, A
					Sequence 54526, A
					Sequence 53, Appli
					Sequence 48, Appli
					Sequence 4401, App
					Sequence 3401, App
					Sequence 1618, App
					Sequence 303359, A
					Sequence 22159, A

85	122	3.3	1342	16	US-10-408-765A-2821	Sequence 2821, App
86	121.5	3.3	828	16	US-10-128-558-234	Sequence 234, App
87	121	3.3	988	15	US-10-424-559-144895	Sequence 144895,
88	120.5	3.2	306	15	US-10-363-493-1389	Sequence 1389, Ap
89	120.5	3.2	481	10	US-09-853-079-211	Sequence 211, App
90	120.5	3.2	481	15	US-10-294-443-211	Sequence 211, App
91	120.5	3.2	481	15	US-10-294-443-327	Sequence 227, App
92	120.5	3.2	1183	17	US-10-485-006A-4	Sequence 4, Appli
93	120.5	3.2	1212	17	US-10-485-006A-2	Sequence 2, Appli
94	120	3.2	293	10	US-09-866-050A-671	Sequence 671, App
95	120	3.2	657	17	US-10-732-923-19475	Sequence 19475, A
96	120	3.2	687	16	US-10-437-963-135760	Sequence 135760,
97	120	3.2	725	16	US-10-719-933-820	Sequence 820, App
98	120	3.2	1235	13	US-10-087-132-363	Sequence 363, App
99	120	3.2	1393	18	US-10-994-106-23	Sequence 23, Appl
100	120	3.2	1393	18	US-10-994-106-47	Sequence 47, Appl

ALIGNMENTS

```

RESULT 1
US-10-650-482-4
; Sequence 4, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN e1f2alpha-S
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN e1f2alpha-S
; FILE REFERENCE: 5986/1L712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-650-482-4

```

		Query Match	100.0%	Score	3710;	DB	16;	Length	698;
		Best Local Similarity	100.0%;	Pred.	No. 2.5e-275;				
		Matches	698;	Conservative	0;	Mismatches	0;	Indels	0;
		Gaps	0;						
Qy	1	METGTHARKRPGRIKSWFRLPLIRSHACSSFFPPSPSRQNGALPERRTRYTKL	60						
Dd	1	METGTHARKRPGRIKSWFRLPLIRSHACSSFFPPSPSRQNGALPERRTRYTKL	60						
Qy	61	LSQLLLPSLFQKLILWSOLSGGLITPRMLDPAASYSALRASGRRESDAPTVOKSLSY	120						
Dd	61	LSQLLLPSLFQKLILWSOLSGGLITPRMLDPAASYSALRASGRRESDAPTVOKSLSY	120						
Qy	121	TAAGLPKTRVWSTLALARGTTPVAIVLRLEVKLKAERALDSAAPTFLLEQQLWGVEL	180						
Dd	121	TAAGLPKTRVWSTLALARGTTPVAIVLRLEVKLKAERALDSAAPTFLLEQQLWGVEL	180						
Qy	181	LPSLQAGLVSHRELDSSSGPLSVQSIGNFKVVSYLLNPSYLDYLQGLRCOSSAGGG	240						
Dd	181	LPSLQAGLVSHRELDSSSGPLSVQSIGNFKVVSYLLNPSYLDYLQGLRCOSSAGGG	240						
Qy	241	QFVGFRTLTPESCVLSDGCHQPQLRAEMSATAWRRCPPLSTEGLPPIHHRMRWLVLQ	300						
Dd	241	QFVGFRTLTPESCVLSDGCHQPQLRAEMSATAWRRCPPLSTEGLPPIHHRMRWLVLQ	300						
Qy	301	PNQGDLPTILDQNGYHSLEEHNLLRMDPOHCTDNPAQAVSPAADRPETEKKPELVIQ	360						
Dd	301	PNQGDLPTILDQNGYHSLEEHNLLRMDPOHCTDNPAQAVSPAADRPETEKKPELVIQ	360						
Qy	361	EVSQSQGSLSLCFELPVEKECEDHTWTDLSDRGESLPVSTRPVCSNKULIDYILGAPS	420						

[illegible]

```

RESULT 2
US-10-650-482-2
; Sequence 2, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron. David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN e
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/1L712-US1
; CURRENT APPLICATION NUMBER: US/10/650.482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-482-2

```

Query Match	59.9%;	Score 2223;	DB 16;	Length 713;
Best Local Similarity	65.0%;	Pred. No. 2.6e-161;		
Matches	456;	Conservative 53;	Mismatches 174;	Indels 18; Gaps 7;
Qy	1	METGTHARCRPGRLGSWFLRPL-RRSHACSSERPPSSRRQNGN-----SALPERRT	54	
Db	1	MEPTGSGSRKRLGRAGFRFWPPFFPRRSQAGSKFETPLGPENSGNPTLLSSAQPTRV	60	
Qy	55	RYMTKLLSOLLALPLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASRGRESDAPTV	114	
Db	61	SYMTKLLSOLLALPLPGLLQKVLWSQLFGWFPTRWLDFAGVYSALRAKGREKPAAPTA	120	
Qy	115	QKLSIYTAAGLFAKTRVWSTLALARGTTPVAVLVLRLEVKLKAQERALDSPAPTFLEEQ	174	
Db	121	QKLSLSLQLD--SSDPSVTSPLDMLKEGIHWQYSPDDLKLELKAKGSSALDPAQAQFLLEEQ	179	
Qy	175	LWGVELLPSSLOAGLYSHRELDSSSGPLTSVQSILGNFKVYSYLLNPYSYLDYLPOLGRQC	234	
Db	180	LWGVELLPSSLOQLRYSNRELGSPPGLNIQRIIDNFSVVSYSYLLNPYSYLDYLPOLGRVC	239	
Qy	235	SSAGGGQFVGFRITLTPBSCYLSLDGCHQPQLRAEMSATARRCPPLSTEGLPETHHRMR	294	
Db	240	NSDGNSEVVGFOILTPBSSCLREDHCHQPQLSAELTPASWQGCPCPLSTEGLPETHHLRMK	299	
Qy	295	WLVFL-OPNQGDLPITLDQNGYHSLEEEHNLRLMDPHQCHTDNPAQVSPAARDP-----	349	
Db	300	RLBFLQOANKGODLPDPDQNGYHSLEEEHSLRLMDPKHCRDNPOTVPVPAAGDTPGNTOE	359	

```

Qy 350 PTEKKPELVIOEV-----SQSPQSSSLFCELPVKEKECEDHTNATDLSRGSLSLPVSTRP 404
Db 360 STEEKIELLTTEVPLALEEESPSCPSSEIPMEKBPCEGRISVVDYSYLEGDLPISARP 419
Qy 405 VCSNKLIDYILGGAPSLDLEASDSSESDEWGPEDDGDGSDGSLSESDVEQDSGLHLWN 464
Db 420 ACSNKLIDYILGGASSDLETSSDPEGDWDEADDGDFDSDSSLSDSLEQDPGGLHLWN 479
Qy 465 SFHSVDPYKPNQFATTIQTAAARIAPRDPDSGTWSGCGV-GSCQEGPLPETPDHSSGE 523
Db 480 SFCSDVPYNPNQFATTIQTAAARIAPRDPDSGTWSGCGV-GSCQEGPLPETPDHSSGE 539
Qy 524 EDDPEPSADEARENKLWNSFCCHSDPYNLNFAPQPSGKWKGRQDSKASSEVTVAFS 583
Db 540 EDDWESSADEABSLLWNSFCNSDDPYNLNFAPQPSGKWKGRQDSKASSEVTVAFS 599
Qy 584 GHHTLLSCKAQLLESQENCGCLGALAGERVTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCKVQLLSQSECPDSVQRDLVSGRTHVKKVTFLEEVTEYIISGDEDRK 659
Qy 644 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLR 684
Db 660 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLQ 700

RESULT 3
US-09-925-299-941
; Sequence 941, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

Query Match 58.2%; Score 2157.5; DB 9; Length 707;
Best Local Similarity 64.2%; Pred. No. 2.7e-156; Indels 17; Gaps 6;
Matches 442; Conservative 54; Mismatches 176; Indels 17; Gaps 6;

Qy 12 PGPLGWSFRLPFLRRSHACSEFPSPSSRQPNQGN-----SALPERRTRYWTKLLSLLA 66
Db 7 PVSRCSGAFQSVRRSQAGSKFPTPLGPNSENGPTLLSSAQPTETRVSVYWKLLSLLA 66
Qy 67 LLPSLFKQLLWSQSLGGLIETRWLDFAASYSALRASGREESDAPTQVKSLSVTAAGLF 126
Db 67 PLPLGLQKVLWSQSLFGGMFTRWLDFAGVYSALRALKAGREKPAAPTAQKSLSLQLD-S 125
Qy 127 AKTRVWSTLALARGTTPVAVLVLEVKLKAQERALDSAAPTFLLEQOLWGVELLPSSLQ 186
Db 126 SDPSVTSPLDWLBEGHWHQVSPFDLKLKELKAKGALSADPAQAFLLEQOLWGVELLPSSLQ 185

```

```

Qy 187 AGLVSHRELDSSSGPLSVQSLGNFKVSVLLNPSYLDYLPOLGLRCSQAGGQFVGPR 246
Db 186 SRLYSNRELSSPSGLNIQRIIDNFVSVYLLNPSYLDYLPOLGLRCSQAGGQFVGPR 245
Qy 247 TLTPESCYLSGDCGHPOLRAEMSAATARRCPPLSTEGLEPEIHHRRMRWLVL- QPNQGG 305
Db 246 TLTPESSCLREDCHQPLXAEELI PKSQWCCPLSTEGLEPEIHHLRMKRLEFLQQAASKG 305
Qy 306 DLFTLDQNGYHSLBEHNLLRMDPOHCTNDPAQAVSPAADRP-----EPTKKPELVIOE 361
Db 306 DXPTPDQNGYHSLBEHSLLRMDPKHCRDNPQTQFVPAAGDI PONTQESTEEKIELTTE 365
Qy 362 V-----SQSPQSSSLFCELPVKEKECEDHTNATDLSRGSLSLPVSTRPVCNKLIDYILG 416
Db 366 VPALLESSESPGCPSEIPEKEPGEGRISVVDYSYLEGDLPI SAR PACSNKLIDYILG 425
Qy 417 GAPSDLEASDSSESDEWGPEDDGDGSDGSLSESDVEQDSGLHLWNSFHSVDPYKPNQ 476
Db 426 GASDLETSSDPEGDWDEADDGDFDSDSSLSDSLEQDPGGLHLWNSFCSVDYPNPQN 485
Qy 477 FTATIQTAAARIAPRDPDSGTWSGCGV-GSCQEGPLPETPDHSSGEEDDWSPSADAE 535
Db 486 FTATIQTAAARIAPRDPDSGTWSGCGV-GSCQEGPLPETPDHSSGEEDDWSPSADAE 545
Qy 536 NLKLWNSFCHSDPYNLNFAPQPSGKWKGRQDSKASSEVTVAFSGHHTLLSCKAQL 595
Db 546 SLKLWNSFCNSDDPYNLNFAPQPSGKWKGRQDSKASSEVTVAFSGHHTLLSCKAQL 605
Qy 596 LESQEDNCPGCLGALAGERVTHIKRKKVTFLEEVTEYIISGDEDRKGPWEEFARDGCR 655
Db 606 LGSQSECPDSVQRDLVSGRTHVKKVTFLEEVTEYIISGDEDRKGPWEEFARDGCR 665
Qy 656 FQKRIQETEDAIGYCLTFEHRERMFNRLR 684
Db 666 FQKRIQETEDAIGYCLTFEHRERMFNRLQ 694

RESULT 4
US-09-925-299-941
; Sequence 941, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

Query Match 58.2%; Score 2157.5; DB 10; Length 707;
Best Local Similarity 64.2%; Pred. No. 2.7e-156; Indels 17; Gaps 6;
Matches 442; Conservative 54; Mismatches 176; Indels 17; Gaps 6;

```

```

Qy 12 PGRILGSMFRLPLRRSHACSEFPSPSRQNGN-----SALPERRTRYWTKLSQLLA 66
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 PVSRCGAFQPSVRSRQAGSKFTPLGPNSENGNPTLLSSAQPETRVSYWTKLSQLLA 66
Qy 67 LLPSLFQKLLWSQLSGGLIPTRWLDPAAASYALRASGREEDSAPTVQKSLSYTAAGLF 126
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 PUGLLOKVLWSQLFGGMFTRWLDFAGVTSALRALGKREKPAAPTAKQSLSSQLD-S 125
Qy 127 AKTRVVTIALARGCTPPAVLVLEVKLKAQERALLDSAAPTFLLEQQLWGVELLPSSLQ 186
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 SDPSTVSLDWLEGIHQVSPDDLUKELKAGSALDPAQAFLLEQQLWGVELLPSSLQ 185
Qy 187 AGLVSHRELDSSSGSLVSQSLGNFKVSYLLNPISYLYLPOLGLRCOSSAGGQFVGFR 246
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
186 SRLYSNRELGGSPGLLIQRIIDNFVSVYLLNPISYLCDFPRLEVSQNSDGNSEVVGQ 245
Qy 247 TLTSPSCYLSGEGCHPQIPARMSATAMRRCPPLSTEGLEPEIHHRMRWLVL-OPNQGQ 305
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 TLTPESSCLREDHCHPQPLXABLIPKSWQCGCPPLSTEGLEPEIHLHMRKLEFLQOASGQ 305
Qy 306 DLPTLDQNGVHSLSEENHLLRMDPHQCTDNPAAQAVSPAADRP-----EPTKEKPELVQE 361
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
306 DXPTPDQNGVHSLSEESHLRMDPKRCDNPTQFVPAAGDIPGWTQSTBEKIELLITE 365
Qy 362 V-----SQSPQSSSLFCBLPVEKECEDHTNATLSDRGESLIPVSPVCSNKLIDYILG 416
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
366 VPLALEESPSEGCPSSEIPMEKEPGEGRISVVDYSYLEGLDIPISARPACSNKLIDYILG 425
Qy 417 GAPSDLEASSDSSEDDWEEBDDDFDSDGSLSESVDQSDSEGLHLWNSFHSVDPYKPN 476
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 GASSDLETSSDPEGDWDDEADDGFDSDSLSDSLSDLEQDPEGLHLWNSFHSVDPYKPN 485
Qy 477 FTATIQTAARTAPRDPSPDSGTSWSSCGV-GSCQSGPLPETPDHSSGEEDDWEPSADRAE 535
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
486 FTATIQTAARIVPEEPSSEKDLKSLENSQSGSLPETPEHSSGEEDDWEPSADRAE 545
Qy 536 NLKLWNSFCHSDPDYNLNFKAPFQPSGKNWKGRODSKASSEVTVAFSGHHTLLSCKAQL 595
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
546 SILKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSESIVASECHTLLSCKVQL 605
Qy 596 LESQEDNCPGGLGELAGERYTHIKRKKVTFLEEVTEYIISGDEDRKGPWEFEPARDGCR 655
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
606 LGSQSECPDSVQRDLVSGRHTHVKKRKKVTFLEEVTEYIISGDEDRKGPWEFEPARDGCR 665
Qy 656 FQKRIQETEVAGYCLAFEREKMFNRLR 684
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
666 FQKRIQETEDAIYCLTFEHRERMFNRLQ 694

```

RESULT 5

```

; Sequence 963, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-963

```

Query Match 14.7%; Score 547; DB 15; Length 153;

Best Local Similarity 76.6%; Pred. No. 9.3e-34;

Matches 105; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

```

Qy 548 DPYNLLNFKAPFQPSGKNWKGRODSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPGCG 607
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 DPYNPLNFKAPFQTSGENEKGCRDSTPSESIVASECHTLLSCKVQLLSQSECECDVS 63
Qy 608 LGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRKGPWEFEPARDGCRFQKRIQETEVAI 667
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 QRDVLSGGRHTHVKKRKKVTFLEEVTEYIISGDEDRKGPWEFEPARDGCRFQKRIQETEDAI 123
Qy 668 GYCLAFEREKMFNRLR 684
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 GYCLTFEHRERMFNRLQ 140

```

RESULT 6

```

US-09-052-753-3
; Sequence 3, Application US/09052753A
; Patent No. US20010014734A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF
; FILE REFERENCE: 0575/51523-B/JPW/JSG
; CURRENT APPLICATION NUMBER: US/09/052,753A
; CURRENT FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Rat
US-09-052-753-3

```

Query Match 5.5%; Score 204; DB 9; Length 657;

Best Local Similarity 21.5%; Pred. No. 1.5e-06;

Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

```

Qy 60 LLSQALLPSLFQKLLWSQLSGGLIPTRWLDPAAASYALRASGREEDSAPTVQKSL 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 LLSPLMGLLSR-----AWSRLRGPEVPEAWL-----AKTVTGADQ 55
Qy 120 YTAAGLPKTRVTVSTLALARG-----GTPVAVLVLRLEVKLKAQERALLDSAPTFLL 171
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
56 IEAALTLTPFVSGNL-LPHGETEESGSP-----EQSQAQRLCLVEAESPP--- 102
Qy 172 EQQLWGVLLPS-SLOAGLVSHRELD-SSSSGSLVSQSLG-----NFKVVSY---LLN 219
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 --ETWGLSNVDEYNAPKPGQDDLEKEMERTAGKATLPAGLQAGDKLGEVVAEEGVAE 160
Qy 220 PSYLDYLPQLGLRCQSSAGGQFVGFRTLTLPESCYLSLEDGCHPQPLRAEMSATAMRRCPP 279
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
161 PAY-----PTSQLEGGPAEN-----EEDGETVKTYQASAASIAPGYKPS 199
Qy 280 LSTEGLEPEIHHRMRWLVLFPNQGGDLPTLDQNGVHSLSEENHLLRMDPHQCTDNPQA 339
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 TPVPFLGAEHQATE---EKGTENKADPNSPSSGSHSRAWEY-YGREKPKQGEAKVE 254
Qy 340 AVSPAADRP---EPTKEKPELVQEVSQSPQSSSLFCBLPVEKECEDHTNATLSDRG 395
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 AHRAGQGHPCRNABAEEGGPETTF--VCTGNAFKAWVYRPGEDTEEDSDSASEDT 312
Qy 396 ESLPVSSTRPVCSNKLIDYILGGAPSDLEASSDSESED-----W-----GE 435
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSASEDTAQTGATPHTSAFLKAWVYRPG 372
Qy 436 EPEDDGFDSGSLSESVDQSDSEGLHLWNSFHSVDPYKPNFATTQTAAARIAPRDPDS 495
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 DTEENSDDLDS--AEEDTAQTGATPHT-SAFKAWVYRPGEDTEEDSDSASEDTAQT 429
Qy 496 GTWSGSGCGVSCQEGPLPETPDHSSGEEDDWEPSADAEENKLWNSFCHSDPYNLLNF 555
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
430 GATPHTSPFLKAWVYRPGEDTEEDTEEDSENVAFGDSETADSSQSPC----- 478
Qy 556 KAPFPQPS----GKNWKGRODSKASSEVTVAFSGHH-----TLLSCKAQLLES-- 598

```

```

Db 479 ---LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAKPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGGLGALAGERYTHIKKKVTFLEEVTEYIIS-----GDEDKGPWEFAR 651
Db 536 RDQDPEIP-----LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAKPLRLQRLRLFKAPT 535
Qy 652 DGRFQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 581 DRSRFAIRIAQAEEKLGPLYLTPDSRARAWARLR 613

RESULT 7
US-10-277-603-3
; Sequence 3, Application US/10277603
; Publication No. US20030219376A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
; FILE REFERENCE: A34608-B-A
; CURRENT APPLICATION NUMBER: US/10/277,603
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/052,753
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Rat
US-10-277-603-3

```

```

Query Match 5.5%; Score 204; DB 15; Length 657;
Best Local Similarity 21.5%; Pred. No. 1.5e-06;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

Qy 60 LLSQALLPSPFQKLLWSQSGGLIPTRWLDPAASYSALRASRGRESDAPTQVKSLS 119
Db 21 LLSPLMGLLR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy 120 YTAAGLFAKTRVSTLALARG-----GTPVAVLRLVLEVKLKAQERALDAAAP-----T 168
Db 56 IERAAALITPTVSGNL-LPHGETESGSP-----EQSQAQRLCLVEAESPP----- 102
Qy 172 EQQLGWVLLPSS-LQAGLVSHRELD-SSSSGGLSVQSLG-----NFKVVS-----LLN 219
Db 103 --ETWGLSNVDEYNAKPGQDDLREKEMERTAGKATLQAPAGLQAGADKRLGEVAREEGVAE 160
Qy 220 PSYLDYLPQLGLRCQSSAGGQFVGRFRTLTPESCYLSEDCGCHPOPLRAEWSATWRRCP 279
Db 161 PAY-----PTSOLEGPAEN-----BEDGETVITYQASAASTIAPGVKPS 199
Qy 280 LSTEGLEPIHRRMRWLVLQPNQGLDPLTDQNGYHSLSEENHLLRMDPQHCTDNPAQ 339
Db 200 TPVPFLGEAEHQATE-----EKGTEKADPNPSGSGSHSRAWEY-YGREKPKQGEAKVE 254
Qy 340 AVSPAARDP-----EPTKKKPELVJQEVSPQSGSLFCELPVEKCEBDHTNATDLDRG 395
Db 255 AHRAGQGHPCRNABEAGGPEPTT--VCTGNAFKAWVYRPGEDTEEDNSDSAEEDT 312
Qy 396 ESLPVSTRPVCNSKLIIDYILGGAESDSEED-----W-----GE 435
Db 313 AQTGATHTSAFLKAWYRPGEDTEEDSDSAEDTAQTGATPHTSAFLKAWVYRPG 372
Qy 436 EPEDGDGDSGLSESVEQSEGLHLWNSFHSVDYKPNQFTATTIQAARIAPRPSDS 495
Db 373 DTEENSGLDS--AEEDTAQTGATHT--SAFLKAWVYRPGEDTEEDNSDLSAEEDTAQT 429
Qy 496 GTSMGSGCGVSGCGEGLPPTPHSSGEEDDWEPSSADEANLKLWNSFCHSDPYNLNLF 555

```

```

Db 430 GATPHTSPFLKAWVYRPGEDTDDTEEDSENVAPEGSETADSSQSPC----- 478
Qy 556 KAPFQPS-----GNWKGRQDSKASSEVTVAFSGHH-----TLLSCKAQLLS-- 598
Db 479 ---LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAKPLRLQRLRLFKAPT 535
Qy 599 --QEDNCPGCGGLGALAGERYTHIKKKVTFLEEVTEYIIS-----GDEDKGPWEFAR 651
Db 536 RDQDPEIP-----LQPORCLPGEKTKGRGEEPLFQVAFYLPCEKPEPWAAKPLRLQRLRLFKAPT 535
Qy 652 DGRFQKRIQETEAIGYCLAFEHREKMFNRLR 684
Db 581 DRSRFAIRIAQAEEKLGPLYLTPDSRARAWARLR 613

RESULT 8
US-10-755-889-90
; Sequence 90, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 90
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-90

```

```

Query Match 5.0%; Score 186.5; DB 16; Length 674;
Best Local Similarity 21.5%; Pred. No. 3.4e-05;
Matches 156; Conservative 73; Mismatches 270; Indels 227; Gaps 32;

Qy 60 LLSQALLPSPFQKLLWSQSGGLIPTRWLDPAASYSALRASRGRESDAPTQVKSLS 118
Db 21 LLSPLMGLLR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
Qy 119 SYTAAGLFAKTRVSTLALARGGTPVAVLRLVLEVKLKAQERALDAAAP-----T 168
Db 56 --VEAGLEGEAR-----TPLA--IPHTPWGRRPGCEAEADSGGPGEDRETGLKT 100
Qy 169 FLLEQQLGWVLLPSS-LQAGLVSHRELDSSSSGGLSVQSLGNFKVSYLLNPSYL----- 223
Db 101 SSSLPEAWG--LLDD--DDGWYGEREATSVPRGSGSQFADGQ---RAPLSPSLLIRTLQ 152
Qy 224 --DYLQGLRLCOSSAGGQFVGRFRTL-----PSCYLSLSEDCGCHPOPLRAEWSATA--- 273
Db 153 GSDKNP--GEEKABEEGVAEEGVNFKSPYPPSHRECCPAVEEEDDEAVKKEAHTSTSA 210
Qy 274 -----WRRCP-----PLSTEGLEPIHRRMRWLVLQPNQGLDPLTDQNGYHSL 320
Db 211 LSPGSKPSTWVSCPGHEENQATDKKTERSKGARKTSVSPRSGSDPRSWEYSGGASEE 270
Qy 321 EEHLLRMDPQHCTDNPAQAVSPAARDPTEKPELVJQEVSPQSGSLFCELPVEKE 380
Db 271 KE-----EKAHEETKGEAA-----PGQSAP-----AQRQLKSWWCQPSDEE 311
Qy 381 CEDHTNATDLDRGSLPVSTRPVCNSKLIIDYILGGAESDSEEDGCEPEDD 440
Db 312 SEVKPLGAAEKQGEAECCPCIPPPSAFLKAWVYRPG---EDTEEEDEE---DEDS 364
Qy 441 GPDSDGSLSESVEQSEGLHL--WNSFHSVDYKPNQFTATTIQ-----TAARIAPRPS 493
Db 365 GSDEEGBEAEASSTTATGVFLKSW-----VIQPGEDTEEDDESDTGSADEREAE 417

```



```
Db 448 -----PARNQGPBEP-----LKGRKVHFSEKVTVHFLAVWAGPAQAARRGPWEQFAR 494
Qy 652 DGRFQKRIQETVEAIGVCLAFEHREKMFNRLR 684
Db 495 DRGRFARRIAQAERQQLGYLTPAFRARAFTLR 527

RESULT 11
US-09-052-753-2
; Sequence 2, Application US/09052753A
; Patent No. US20010014734A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF
; FILE REFERENCE: 0575/51523-B/JPM/JSG
; CURRENT APPLICATION NUMBER: US/09/052,753A
; CURRENT FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Rat
US-09-052-753-2

Query Match 4.6%; Score 171.5; DB 9; Length 590;
Best Local Similarity 18.1%; Pred. No. 0.00039;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSDGCHPOPLRAEMSAATAWRCPP-----STEGLEPEIHRR 292
Db 103 PETLGLSDDDKQGDGPREGRAHTAGLPILLSFGLQSAKSLGVEVAGVETELAYPT 162
Qy 293 MRWLVLQPNQGDLPPLDQNG-----YHSLSEEH 323
Db 163 SHW-----EGCPSEEDGETVKAFRASADSPGHKSSTSVYCPGAEHQATEKQ 213
Qy 324 NLRMDP-----QHCTDNPAQVSPAADRPEP-----TEKKPELV 358
Db 214 TENKADPPSPSGSHRAWEYCSKQGEA-----DPEPHRAGKYQLCQNAEAESEBEAK 267
Qy 359 IQEVSQSPQSSSL--FCELPVEKECEDHTNATDLSRGESLPVSTRPVCNKLIDYILG 416
Db 268 VSSLVSNGNAFLKAWYRFGEDTDDDDSDWGSABEEGKALSPTSPE-----HDFLKA 322
Qy 417 GAPSDLEASDSESDGWERPEDDGFDSGSLSDVEQDSEGLHLWNSFHSVDPPYKPN 476
Db 323 WYTRPGEDTDDDDSDWGSABE-----EGKALSPTSPEHDFLKA-----VYRGE 369
Qy 477 FTATIQTAAARIAPRDPDSGTSWSGCGVSGCQEGPLPET-----516
Db 370 DT-----EDDQSDWGSABEKGGLAQTATPHTSAFLKTWVCCPGEDTDDDC 416
Qy 517 -----PDHSSGEEDWBS-----ADEAENKLWNSFCHSEDPYNLLNFKAPFPQSGK 564
Db 417 EVVVPEDSEADPKSPSHEAQGLPGEQTEGL-----VEAHSFLQVAFYLPGEKPA 470
Qy 565 NWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPCGGLGEALAGERYTHIRKK 624
Db 471 PW-----TAPKPLRLQRLTLRLTPQTQ-----DQDPETP-----LRARK 505
Qy 625 VTFLSEVTEYIIS-----GDEDRKGPWEFARDGCRFQKRIQETVEAIGVCLAFEHREK 679
Db 506 VFHSENVTVHFLAVWAGPAQAARRGPWEQLARSRFARRIAQAERKGLGYLTPAFRARA 565

Qy 680 FNRLRIESKDLLL 692
Db 566 WARLGNPSLPLAL 578

RESULT 12
US-10-277-603-2
; Sequence 2, Application US/10277603
```

```
; Publication No. US20030219376A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
; FILE REFERENCE: A34608-B-A
; CURRENT APPLICATION NUMBER: US/10/277,603
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/052,753
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Rat
US-10-277-603-2

Query Match 4.6%; Score 171.5; DB 15; Length 590;
Best Local Similarity 18.1%; Pred. No. 0.00039;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSDGCHPOPLRAEMSAATAWRCPP-----STEGLEPEIHRR 292
Db 103 PETLGLSDDDKQGDGPREGRAHTAGLPILLSFGLQSAKSLGVEVAGVETELAYPT 162
Qy 293 MRWLVLQPNQGDLPPLDQNG-----YHSLSEEH 323
Db 163 SHW-----EGCPSEEDGETVKAFRASADSPGHKSSTSVYCPGAEHQATEKQ 213
Qy 324 NLRMDP-----QHCTDNPAQVSPAADRPEP-----TEKKPELV 358
Db 214 TENKADPPSPSGSHRAWEYCSKQGEA-----DPEPHRAGKYQLCQNAEAESEBEAK 267
Qy 359 IQEVSQSPQSSSL--FCELPVEKECEDHTNATDLSRGESLPVSTRPVCNKLIDYILG 416
Db 268 VSSLVSNGNAFLKAWYRFGEDTDDDDSDWGSABEEGKALSPTSPE-----HDFLKA 322
Qy 417 GAPSDLEASDSESDGWERPEDDGFDSGSLSDVEQDSEGLHLWNSFHSVDPPYKPN 476
Db 323 WYTRPGEDTDDDDSDWGSABE-----EGKALSPTSPEHDFLKA-----VYRGE 369
Qy 477 FTATIQTAAARIAPRDPDSGTSWSGCGVSGCQEGPLPET-----516
Db 370 DT-----EDDQSDWGSABEKGGLAQTATPHTSAFLKTWVCCPGEDTDDDC 416
Qy 517 -----PDHSSGEEDWBS-----ADEAENKLWNSFCHSEDPYNLLNFKAPFPQSGK 564
Db 417 EVVVPEDSEADPKSPSHEAQGLPGEQTEGL-----VEAHSFLQVAFYLPGEKPA 470
Qy 565 NWKGRQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPCGGLGEALAGERYTHIRKK 624
Db 471 PW-----TAPKPLRLQRLTLRLTPQTQ-----DQDPETP-----LRARK 505
Qy 625 VTFLSEVTEYIIS-----GDEDRKGPWEFARDGCRFQKRIQETVEAIGVCLAFEHREK 679
Db 506 VFHSENVTVHFLAVWAGPAQAARRGPWEQLARSRFARRIAQAERKGLGYLTPAFRARA 565

Qy 680 FNRLRIESKDLLL 692
Db 566 WARLGNPSLPLAL 578

RESULT 13
US-10-425-114-38378
; Sequence 38378, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
```

```

; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 38378
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700094420_FLI.pep
US-10-425-114-38378

```

Query Match	4.1%;	Score 152.5;	DB 15;	Length 612;
Best Local Similarity	21.5%;	Pred. No. 0.012;		
Matches 90;	Conservative 59;	Mismatches 133;	Indels 137;	Gaps 22;
Qy	307	LPTLDQ----	DNGVHSL-----	EEBHNLLRMD---POHCTDNPAQAVSPAADRPETKECP 355
Db	274	LPAMSKLWADSDWRSLSKVQWDEFSILRPDRISFEV--EPLDRAHPQSPQPIRAKRP 331		
Qy	356	ELVIQE--VSQSPQSSLFCELPVEKECEDHTNATLSDRGSBLPVRPVCNSKLIDY 413		
Db	332	RPPASPQWSELPSGFGWL-KSPIESSCTLSFSPQRARELPFGIPITSTLSSSN----- 385		
Qy	414	ILGAPSDLEASSDSBEDWCEEBDDGDFSDGSLSDSDVQDSEGLHLWNSFHSDVPYK 473		
Db	386	-----VSFNKNEP-----	SMLTSQ-----	FYWSARHT---R 409
Qy	474	PQNFTATIQTAARIAPRDPDSGTSWSGSC---GVGSC--QEGPLPETPDHSSGEEDWE 528		
Db	410	ADSCAASTNTVIEKKQEPS-----	SGGCRLLFGINICSAEEVLPEVTAFGVG----	YE 459
Qy	529	PSADEAENKLWNSFCHSDPYNNLKFAPPQPSGKWKGQDS-KASSEVTVAFSGHHT 587		
Db	460	QTAASVE-----	LNSDKLSQPSDVN---NSDALAASSERSPLSQSRQ 499	
Qy	588	LLSCKAQLLESQEDNCPGCLGEA-----	LAGERYTHLKRKKVT 626	
Db	500	VRSCVKVIMQ-----	GMAVGRAVDITKLSGYSDLCQKLEEMFDIQELGSLTKLKKRWI 552	
Qy	627	FLEEVTRYISGDBRKPWEFPARDGCRFKRIQETEAIVGYCIAFHEHKMFNRURI 685		
Db	553	FTDDEDDMLVGDG---PWDEF---CRMVRI-----	YITYVEAAKLTASKSL 595	

```

RESULT 14
US-10-425-115-207028
US-10-425-115-207028, Application US/10425115
; Sequence 207028, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 207028
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(661)
; OTHER INFORMATION: unsure at all Xaa locations

```

```
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577_120398C.1.pep  
US-10-425-115-207028
```

```
Query Match          4.1%; Score 152.5; DB 16; Length 661;  
Best Local Similarity 21.5%; Pred. No. 0.013;  
Matches    90; Conservative   59; Mismatches 133; Indels 137; Gaps 22;
```

```
Qy 307 LPTLDQ-----DNGVHSL-----EEHNLLRMD----PQHCTNPAQAQVSPAADRPEPETEKP 355  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 323 LPAMSKLWADSWRSILKVDWPESIIPLDRISLPWEV--EPLDAANPOSPQPPLRAIKRP 380  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 356 ELVIQE--VSQSPOGSSILFCELPVEKECEDHTNATDLSDRGESSIPVSTRPVCSNKLI DY 413  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 381 RPPASPCMWSELPSGGFLW-KSPIESCSTLSFSFPQRARELFPIPISTLSSSN----- 434  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 414 ILGAPSDLLEASSDSSESGDWGEEDPDGDGSLSGSDVEDQSEGHLWLNSFFHSVDYK 473  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 435 -----VSFNKNPE-----SMLTSQ-----PYWSARHT---r R 458  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 474 PQNFATTIQTAAARIAPDPDSGTSWGSC-----GVGSC--QEGPLPETPDHSSGEDDWE 528  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 459 ADSCAAGTNTVIIKKQEPS-----SGCKLFGINICSAEEVLPEVTARGVG-----YE 508  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 529 PSADAENKLNLWNSFCHESEDYPYNLLNPAPPQPSGKNWKGRQDS-KASSEVTPAFSGHHT 587  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 509 QTAASVE-----LNSDKLSOPSDVN---NSDALAASSERSPLESQSRQ 548  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 588 LLSCKAQLLESQDNCPGCGLGEA-----LAGERYTHIKRKKVY 626  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 549 VRSETCKVMIO-----GMAVGRAVDUTKLGYSDLCOKLEEMFDIQELGSTLUKKRWI 601  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Qy 627 FLEEVEYYISGDBDKPWPEEFARDGCRFKRIQETAEVAIGYCIAFEHRKFMRNLRI 685  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :  
Db 602 FTDDDDMMIVGLD-----PWDER-----CRMWRKI-----YLYTVBEAKLKSKSL 644  
||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
```

```

RESULT 15
US-10-425-114-66899
; Sequence 66899, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 66899
; LENGTH: 830
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-2MFLMO17068A01_FLI pep
US-10-425-114-66899

```

	Query Match	4.1%	Score 152.5	DB 15	Length 830
	Best Local Similarity	21.5%	Pred. No. 0.018		
	Matches 90	Conservative	Mismatches 133	Indels 137	Gaps 22
Qy	307	LPTLDQ----	DNGVHSL-----	EEHNLLRMD----	POHCTDNPQAQVSPAADRPETKKP 355
Db	492	LPAMSKSLWADSWRS	SLKVMDFSSILRDP	RSIPWVEV--	EPDLAAVPSQSPQLRAKRP 549
Qy	356	ELVIQE--	VQSQSGSGLFCELP	VEKECEBDHTNATL	DSRGESLPVSTRPVCNKILDY 413
Db	550	RPASPQMWSELPSG	FGIWM-KSIPESCTLS	FSFEPQARLEFSP	IPSTLSSSSN----- 603

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:40:09 ; Search time 178.581 Seconds
(without alignments)
4565.254 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHARKRPGRLGSWF.....MFNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6959266 seqs, 1168006243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- Pending Patents AA Main:*
- 1: /cgn2_6/ptodata/1/paa/PCRTUS_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
 - 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
 - 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
 - 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
 - 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
 - 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
 - 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
 - 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
 - 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
 - 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
 - 31: /cgn2_6/ptodata/1/paa/US105_COMB.pep.*
 - 32: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
 - 33: /cgn2_6/ptodata/1/paa/US107_COMB.pep.*
 - 34: /cgn2_6/ptodata/1/paa/US108_COMB.pep.*
 - 35: /cgn2_6/ptodata/1/paa/US109_COMB.pep.*
 - 36: /cgn2_6/ptodata/1/paa/US110_COMB.pep.*
 - 37: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3710	100.0	698	32	US-10-650-482-4	Sequence 4, Appl1
2	2223	59.9	713	32	US-10-650-482-2	Sequence 2, Appl1
3	2218	59.8	713	37	US-10-170-205E-23916	Sequence 23916, A
4	2218	59.8	713	37	US-60-452-680-17655	Sequence 17655, A
5	2213	59.6	713	37	US-09-629-469A-18965	Sequence 18965, A
6	2213	59.6	713	35	US-10-917-503-18965	Sequence 18965, A
7	2208	59.5	720	22	US-09-757-028-2348	Sequence 2348, Ap
8	2208	59.5	720	28	US-10-222-911-2348	Sequence 2348, Ap
9	2157.5	58.2	707	1	PCT-US00-05883-941	Sequence 941, App
10	2157.5	58.2	707	24	US-09-925-299-941	Sequence 941, App
11	547	14.7	153	1	PCT-US00-35017A-963	Sequence 963, App
12	547	14.7	153	28	US-10-296-115-963	Sequence 15743, A
13	357	9.6	101	1	PCT-US01-14827-15743	Sequence 15743, A
14	307	8.3	67	37	US-60-196-718-7356	Sequence 7356, Ap
15	304	8.2	73	37	US-60-177-571-4491	Sequence 4491, Ap
16	304	8.2	73	37	US-60-162-247-4518	Sequence 4518, Ap
17	304	8.2	73	37	US-60-169-840-8186	Sequence 8186, Ap
18	204	5.5	657	12	US-08-893-852-3	Sequence 3, Appl1
19	204	5.5	657	28	US-10-274-666-3	Sequence 3, Appl1
20	204	5.5	657	28	US-10-277-603-3	Sequence 1, Appl1
21	191.5	5.2	674	12	US-08-893-852-1	Sequence 905, App
22	191.5	5.2	697	34	US-10-821-234-905	Sequence 18860, A
23	189.5	5.1	674	27	US-10-170-205E-18860	Sequence 549, App
24	189.5	5.1	674	28	US-10-219-051B-549	Sequence 553, App
25	189.5	5.1	674	28	US-10-219-051B-553	Sequence 10552, A
26	189.5	5.1	674	28	US-10-219-051B-10552	Sequence 11494, A
27	189.5	5.1	674	37	US-60-443-566-4148	Sequence 4148, Ap
28	189.5	5.1	674	37	US-60-452-680-23269	Sequence 23269, A
29	189.5	5.1	674	37	US-60-453-050-14551	Sequence 14551, A
30	189.5	5.1	674	37	US-60-453-135-14551	Sequence 14551, A
31	189.5	5.1	674	37	US-60-455-444-7838	Sequence 7838, Ap
32	189.5	5.1	674	37	US-60-465-241-7838	Sequence 7838, Ap
33	189.5	5.1	674	37	US-60-466-412-14551	Sequence 14551, A
34	186.5	5.0	674	20	US-09-629-469A-11494	Sequence 11494, A
35	186.5	5.0	674	33	US-10-755-889-90	Sequence 90, Appl1
36	186.5	5.0	674	35	US-10-917-503-11494	Sequence 11494, A
37	186.5	5.0	674	37	US-60-440-068-90	Sequence 90, Appl1
38	186.5	5.0	674	37	US-60-469-757-90	Sequence 90, Appl1
39	182.5	4.9	578	16	US-09-236-772-2	Sequence 2, Appl1
40	182.5	4.9	578	28	US-10-274-666-7	Sequence 7, Appl1
41	182.5	4.9	578	28	US-10-277-603-7	Sequence 7, Appl1
42	171.5	4.6	590	12	US-08-893-852-4	Sequence 4, Appl1
43	171.5	4.6	590	28	US-10-274-666-2	Sequence 2, Appl1
44	171.5	4.6	590	28	US-10-277-603-2	Sequence 2, Appl1
45	156	4.2	627	21	US-09-724-676-55843	Sequence 55843, A
46	156	4.2	627	21	US-09-724-676A-55843	Sequence 31529, A
47	152.5	4.1	612	27	US-10-155-881-31529	Sequence 38378, A
48	152.5	4.1	612	30	US-10-425-114-38378	Sequence 38378, A
49	152.5	4.1	612	30	US-10-425-114A-38378	Sequence 17030, A
50	152.5	4.1	612	37	US-60-324-109-17030	Sequence 207028, A
51	152.5	4.1	661	30	US-10-425-115-207028	Sequence 5980, Ap
52	152.5	4.1	661	30	US-10-438-246-5980	Sequence 58227, A
53	152.5	4.1	830	28	US-10-219-999-58227	Sequence 66899, A
54	152.5	4.1	830	30	US-10-425-114-66899	Sequence 66899, A
55	152.5	4.1	830	30	US-10-425-114A-66899	Sequence 27894, A
56	152.5	4.1	830	37	US-60-324-109-27894	Sequence 7333, Ap
57	152.5	4.1	834	27	US-10-155-881-7333	Sequence 152, App
58	149.5	4.0	923	1	PCT-US02-10780-152	Sequence 1354, Ap
59	149.5	4.0	923	27	US-10-114-270-152	Sequence 1354, Ap
60	149	4.0	1365	1	PCT-US03-40978-1354	Sequence 704, App
61	149	4.0	1365	33	US-10-741-600-1354	Sequence 439, App
62	149	4.0	1365	33	US-10-796-307-704	Sequence 13533, A
63	149	4.0	1365	35	US-10-995-561-867	Sequence 13533, A
64	149	4.0	1365	37	US-60-568-219-439	Sequence 11065, A
65	148.5	4.0	676	20	US-09-614-150-13533	Sequence 13533, A
66	148.5	4.0	676	20	US-09-614-150A-13533	Sequence 11065, A
67	148.5	4.0	676	22	US-09-791-537-11065	Sequence 13595, A
68	148.5	4.0	676	37	US-60-167-217-13595	Sequence 11069, A
69	148.5	4.0	676	37	US-60-173-464-11069	

70 148.5 4.0 676 37 US-60-191-637-13574 Sequence 13574, A
71 148.5 4.0 676 37 US-60-191-681-10700 Sequence 10700, A
72 147.5 4.0 725 20 US-09-629-469A-17386 Sequence 17386, A
73 147.5 4.0 725 35 US-10-917-503-17386 Sequence 17386, A
74 147.5 4.0 728 27 US-10-170-205B-22564 Sequence 22564, A
75 147.5 4.0 728 37 US-60-452-680-17245 Sequence 17245, A
76 147.5 4.0 728 37 US-60-453-050-10486 Sequence 10486, A
77 147.5 4.0 728 37 US-60-453-135-10486 Sequence 10486, A
78 147.5 4.0 728 37 US-60-466-412-10486 Sequence 10486, A
79 147.5 4.0 835 27 US-10-170-205E-23688 Sequence 23688, A
80 147.5 4.0 835 33 US-10-755-889-36 Sequence 36, Appl
81 147.5 4.0 835 37 US-60-440-068-36 Sequence 36, Appl
82 147.5 4.0 835 37 US-60-452-680-17244 Sequence 17244, A
83 147.5 4.0 835 37 US-60-453-050-10485 Sequence 10485, A
84 147.5 4.0 835 37 US-60-453-135-10485 Sequence 10485, A
85 147.5 4.0 835 37 US-60-466-412-10485 Sequence 10485, A
86 147.5 4.0 835 37 US-60-469-757-36 Sequence 36, Appl
87 144 3.9 867 1 PCT-US01-29288A-130 Sequence 130, Appl
88 144 3.9 867 1 PCT-US01-29288A-130 Sequence 130, Appl
89 144 3.9 867 18 US-09-487-558-98 Sequence 98, Appl
90 144 3.9 867 22 US-09-487-558B-98 Sequence 98, Appl
91 144 3.9 867 22 US-09-791-537-92178 Sequence 92178, A
92 144 3.9 867 23 US-09-801-368-98 Sequence 98, Appl
93 144 3.9 867 27 US-10-149-310-130 Sequence 130, Appl
94 142.5 3.8 917 1 PCT-US04-24424-1542 Sequence 1542, Ap
95 142.5 3.8 917 37 US-60-490-890-1542 Sequence 1542, Ap
96 141 3.8 917 1 PCT-US02-26322-15 Sequence 15, Appl
97 141 3.8 917 30 US-10-487-092-15 Sequence 15, Appl
98 141 3.8 1198 33 US-10-777-288A-3349 Sequence 3349, Ap
99 139.5 3.8 883 1 PCT-US02-22539-6 Sequence 6, Appl
100 139.5 3.8 883 1 PCT-US03-17500-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-10-650-482-4
; Sequence 4, Application US/10650482
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-650-482-4

Query Match 100.0%; Score 3710; DB 32; Length 698;
Best Local Similarity 100.0%; Pred. No. 1e-313;
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 METGTHARKPGPRLGSWFRLPFLRRSHACSSBPPPPSSRQNPNCNSALPERRTRYWTKL 60
Db 1 METGTHARKPGPRLGSWFRLPFLRRSHACSSBPPPPSSRQNPNCNSALPERRTRYWTKL 60
Qy 61 LSQALLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASGRBESDAPTQKLSY 120
Db 61 LSQALLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASGRBESDAPTQKLSY 120
Qy 121 TAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQQLWGVEL 180
Db 121 TAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQQLWGVEL 180

181 LPSSLQAGLVSHRELSSSSGPLSVQSLGNFKVSVYLLNPSYLDYLPQLGLRCOSSAGG 240
181 LPSSLQAGLVSHRELSSSSGPLSVQSLGNFKVSVYLLNPSYLDYLPQLGLRCOSSAGG 240
241 QFVGFRLLTPESCYSYSDGCHPOPLRAEMGATARRCPPLSTEGLPBIHRRMRWLVLQ 300
241 QFVGFRLLTPESCYSYSDGCHPOPLRAEMGATARRCPPLSTEGLPBIHRRMRWLVLQ 300
301 PNOGQDLPTLDQNGVHSLFEEHNLRLMDPQHCTDPAQAVSPAADPEPEKPELVQ 360
301 PNOGQDLPTLDQNGVHSLFEEHNLRLMDPQHCTDPAQAVSPAADPEPEKPELVQ 360
361 EVSPOGSSSLFCELPVEKECEDHTNATDLSRGESELPVSTRPVCSNKLIDYILGAPS 420
361 EVSPOGSSSLFCELPVEKECEDHTNATDLSRGESELPVSTRPVCSNKLIDYILGAPS 420
421 DLRASSDSSEDEWGEPEDDGDFDGSLSDEVEQDSEGLHLMNSFHSVDPYKQNFAT 480
421 DLRASSDSSEDEWGEPEDDGDFDGSLSDEVEQDSEGLHLMNSFHSVDPYKQNFAT 480
481 IOTAARTAPRDPDSGTSWSSCGVSCQSGPLPETPDHSSGREDDEWPSADEAENIKLW 540
481 IOTAARTAPRDPDSGTSWSSCGVSCQSGPLPETPDHSSGREDDEWPSADEAENIKLW 540
541 NSFCHSEDPYNLLNFKAPFQPSGKNWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQE 600
541 NSFCHSEDPYNLLNFKAPFQPSGKNWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQE 600
601 DNPFGCGLGALAGERYTHIKRKVTPLEVTYISGDDRKGPWEFARDGCRFOKRI 660
601 DNPFGCGLGALAGERYTHIKRKVTPLEVTYISGDDRKGPWEFARDGCRFOKRI 660
661 QETEVAIGYCLAFEHREKMFNRLRIESKOLLVSNVKK 698
661 QETEVAIGYCLAFEHREKMFNRLRIESKOLLVSNVKK 698

RESULT 2

US-10-650-482-2
; Sequence 2, Application US/10650482
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-482-2

Query Match 59.9%; Score 2223; DB 32; Length 713;
Best Local Similarity 65.0%; Pred. No. 7.8e-184;
Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;
Qy 1 METGTHARKPGPRLGSWFRLPFLRRSHACSSBPPPPSSRQNPNCNSALPERRTRYWTKL 54
Db 1 METGTHARKPGPRLGSWFRLPFLRRSHACSSBPPPPSSRQNPNCNSALPERRTRYWTKL 54
Qy 55 RYWTLLSQALLPSLFQKLLWSQSGGLIPTRWLDFAASYSALRASGRBESDAPTQ 114
Db 61 SYWTLLSQALLPSLFQKLLWSQSGGLIPTRWLDFAAGVTSALRALGREGKPAAPTA 120
115 QKLSYTAAGLFAKTRVVTALARGGTPVAVLVLRLEVKLKAERALDSAAPTFLLEQ 174
121 QKLSLSQLD-SSDPSVTSPLDWLEGIHWQYSPDPLKLEKAGKSALDPAQAFLEEQ 179

Qy	175	LMGWELLPSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQ	234
Db	180	LMGWELLPSLQSRLYSNRELGGSPGGLNIQRIDNFSVSYLLNPSYLDLCPRLVSVQ	239
Qy	235	SSAGGGQVGRPTLTPESCYLEDCCHPOPLRAEMSAATAWRCRCPPLSTEGLEPIHHRRM	294
Db	240	NSDGNSEVVGFTLTPTESSCLURDHCHPOPLASAEIIPASWQCOPPLSTEGLEPIHHLRMK	299
Qy	295	WLVL-OPNQOQDLPTLDDQNGYHSLBEHNLRLMDPOHCTDNPAQAVSPAADRP----	349
Db	300	RLEFLQANKGODLPTPDODNGYHSLBEHSLRLMDPKHCRDNPQFVPAAGDIPGNTQE	359
Qy	350	PTEKKPELVIQEV-----SQSPQGSLSLFCELPVKEKECEDHTNATDLSDRGESLPVSTRP	404
Db	360	STEBKIELLTTEVPLALBEESPSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLPIARP	419
Qy	405	VCSNKLIDYIILGGAPSDLEASSDSSESDGWBEPDDGFDSDGSLSESVDQSEGLHLWN	464
Db	420	ACSNKLIDYIILGGASSDLTSSDPGEHDWEAEDDGFDSLSLSDLEQDEPGLHLWN	479
Qy	465	SFHSVDYPKQNFATTIQTAAIRAPRDPDSGTSWSGCGV-GSQEQGLPETPDHSSGE	523
Db	480	SFCSVDYPNQNFATTIQTAAIRIVPEEPSDEKDLGSKDLENSQSGSLPETPEHSSGE	539
Qy	524	EDDWEPSADAEENLKLWNSFCHSEDPYNLLNFKAPQPSGKKWKGRQDSKASSEVTVAFS	583
Db	540	EDDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKCRCDSKTPSEISVAIS	599
Qy	584	GHTLLSCKAQLLESOBONCPGCLGEALAGERYTHIKRKVKYTFLEEVTEYYISGDDEBK	643
Db	600	ECHTLLSCKVQLLSQSESECPSVQRDVLSGRHTHVRKKVTFLEEVTEYYISGDDEBK	659
Qy	644	GPWEEFARDGCRFOKRIQETVEIVAIGCYCLAFEHREKQFNRLR	684
Db	660	GPWEEFARDGCRFOKRIQETDAIGCYCLTFEHRERMFNRLO	700

RESULT 3
 US-10-170-205E-23916
 ; Sequence 23916, Application US/10170205E
 ; GENERAL INFORMATION:
 ; APPLICANT: ADAMS, Mark
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
 ; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
 ; FILE REFERENCE: CLO01381
 ; CURRENT APPLICATION NUMBER: US/10/170,205E
 ; CURRENT FILING DATE: 2002-06-13
 ; NUMBER OF SEQ ID NOS: 40312
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23916
 ; LENGTH: 713
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-170-205E-23916

```

RESULT 4
US-60-452-680-17655
; Sequence 17655, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17655
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-17655

```

	Query Match	59.8*	Score 2218;	DB 37;	Length 713;	
	Best Local Similarity	64.9*;	Pred. No. 2.1e-103;			
	Matches 455;	Conservative 54;	Mismatches 174;	Indels 18;	Gaps 7;	
Qy	1	METGTHRAKRP	GGPRLGSWPLRPLF-	RRSHACSSSEFP	PPPSRSQNPGN-----SALPERRT	54
Db	1	MEPGTGSRKRLGPRAGFRFWPPP	PPRRSQSGSKFFTP	LGPENSGNPTLLSSAQPETRV	60	
Qy	55	RYWTKLSQLALLPSLFQKLLWSQSGLLI	PTRWLDFAAYS	SALRASRGRESDAPTV	114	
Db	61	SYWTKLLSQLALPLPGLLKVL	IWSQFGGMFTRWLD	FAGVYSAIRALKGRKPAAPTA	120	
Qy	115	QKSISYTAAGLF	FAKTRVVSTILAR	GTTPAVLVLRLEVKLQA	GERALDSAAPFLLEQQ	174
Db	121	QKSUSSIQLD--SSDPSTVSPDLWUEEGH	HWOQVSPPDLKLELUK	AGSAUDPAAQAFLEBQQ	179	
Qy	175	LMGWYELLFPSSI	LAQLGVSHREDLSSSGPLFS	VQSIGNFKWSYLLNP	SYLDLPOLGLRCQ	234

```

Db 180 LWGVLLPSSLSQRLSYNRELSSPSGPNLQIRIDNFVSVYLLNPSYLDLCPRLVSYQ 239
Qy 235 SSAGGGQFVGRPTLTPESCYLSEDCCHQPLPAENSAATAWRCCPLSTEGLPETHHRRMR 294
Db 240 NSDGNSEVVGFOTLTPESCLREDCCHQPLPAENSAATAWRCCPLSTEGLPETHHRRMR 299
Qy 295 WLVLFL-QPNQGDPLTLQDNGYHSLSEBHNLRMDPQCHTDNPAQAVSPAADRP---E 349
Db 300 RLEFLQASKGQDLTPDQDNGYHSLSEBHNLRMDPQCHTDNPAQAVSPAADRP---E 359
Qy 350 PTEKKPELVIOEV-----SQSQSGSLFCELPVEKECEDHTNATDLSDRGESLPVSTRP 404
Db 360 STEEKIELLTTEVPALBEESPSECPSPMEKEGEGRISVVDVSYLSEGLDLPISARP 419
Qy 405 VCSNKLIDYILGGASDLSESDWGEDWDEAEEDDGFDSLSSESDVEODSEGLHLWN 464
Db 420 ACSNKLIDYILGGASDLSESDWGEDWDEAEEDDGFDSLSSESDVEODSEGLHLWN 479
Qy 465 SFHSDVPYKPNFTATTIQTAAARIAPRDPDSGTSWSGCGV--GSCQEGPLPETPDHSSGE 523
Db 480 SFCSDVPYKPNFTATTIQTAAARIAPRDPDSGTSWSGCGV--GSCQEGPLPETPDHSSGE 539
Qy 524 EDDWESSADEARNLKNWSPCHSEDPYNLNFKAPFQPSGKNWKGRODSKASSEVTVAFS 583
Db 540 EDDWESSADEARNLKNWSPCHSEDPYNLNFKAPFQPSGKNWKGRODSKASSEVTVAFS 599
Qy 584 GHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTVYISGDEDRK 643
Db 600 ECHTLSCVKQLLGSQSECPDSVQRDLVSGRHTHVKKKVTFLFEEVTVYISGDEDRK 659

```

RESULT 5

```

US-09-629-469A-18965
; Sequence 18965, Application US/09629469A
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629,469A
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PR1

```

```

; ORGANISM: Homo sapiens
US-09-629-469A-18965

Query Match      59.6%; Score 2213; DB 20; Length 713;
Best Local Similarity 64.8%; Pred. No. 5.8e-183;
Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METCTHARARPGRLGSLWFLPFL-RRSHACSEPPPPSSSRONPGN-----SALPERRT 54
Db 1 MEPCTGSRARLGPAGRFWPPFFRRSQAGSKFTPLGPPENSGNPTLLSSAQPETRV 60
Qy 55 RYWTKLSQLALLPSLQKLLWSQLSGGLIPTRWLDFAASYSALRASGRESDAPTV 114
Db 61 SYWTKLSQLALLPSLQKLLWSQLSGGLIPTRWLDFAAGVTSALRAALGREKPAAPTA 120
Qy 115 QKSLSYTAAGLFAKTRVVTIALARGTTPAVLVLRLEVKLKAQERALDLSAATFLLEQQ 174
Db 121 QKSLSSQLD--SSDPSVTSPLDWLEGIHWQYSPDPLKBLKAGKALSADFAAQFLLEQQ 179
Qy 175 LWGVLLPSSLSQAGLVSHRELDSSSGPLSVQSLGNFKVSVYLLNPSYLDLPOLGRLCQ 234
Db 180 LWGVLLPSSLSQAGLVSHRELDSSSGPLSVQSLGNFKVSVYLLNPSYLDLPOLGRLCQ 239
Qy 235 SSAGGGQFVGRPTLTPESCYLSEDCCHQPLPAENSAATAWRCCPLSTEGLPETHHRRMR 294
Db 240 NSDGNSEVVGFOTLTPESCLREDCCHQPLPAENSAATAWRCCPLSTEGLPETHHRRMR 299
Qy 295 WLVLFL-QPNQGDPLTLQDNGYHSLSEBHNLRMDPQCHTDNPAQAVSPAADRP---E 349
Db 300 RLEFLQASKGQDLTPDQDNGYHSLSEBHNLRMDPQCHTDNPAQAVSPAADRP---E 359
Qy 350 PTEKKPELVIOEV-----SQSQSGSLFCELPVEKECEDHTNATDLSDRGESLPVSTRP 404
Db 360 STEEKIELLTTEVPALBEESPSECPSPMEKEGEGRISVVDVSYLSEGLDLPISARP 419
Qy 405 VCSNKLIDYILGGASDLSESDWGEDWDEAEEDDGFDSLSSESDVEODSEGLHLWN 464
Db 420 ACSNKLIDYILGGASDLSESDWGEDWDEAEEDDGFDSLSSESDVEODSEGLHLWN 479
Qy 465 SFHSDVPYKPNFTATTIQTAAARIAPRDPDSGTSWSGCGV--GSCQEGPLPETPDHSSGE 523
Db 480 SFCSDVPYKPNFTATTIQTAAARIAPRDPDSGTSWSGCGV--GSCQEGPLPETPDHSSGE 539
Qy 524 EDDWESSADEARNLKNWSPCHSEDPYNLNFKAPFQPSGKNWKGRODSKASSEVTVAFS 583
Db 540 EDDWESSADEARNLKNWSPCHSEDPYNLNFKAPFQPSGKNWKGRODSKASSEVTVAFS 599
Qy 584 GHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTVYISGDEDRK 643
Db 600 ECHTLSCVKQLLGSQSECPDSVQRDLVSGRHTHVKKKVTFLFEEVTVYISGDEDRK 659
Qy 644 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLR 684
Db 660 GPWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLQ 700

```

RESULT 6

```

US-10-917-503-18965
; Sequence 18965, Application US/10917503
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123

```


; CURRENT APPLICATION NUMBER: US/10/917,503									
; CURRENT FILING DATE: 2004-08-13									
; PRIOR APPLICATION NUMBER: US/09/629,469									
; PRIOR FILING DATE: 2000-07-28									
; PRIOR APPLICATION NUMBER: JP 1999-248036									
; PRIOR FILING DATE: 1999-07-29									
; PRIOR APPLICATION NUMBER: JP 1999-300253									
; PRIOR FILING DATE: 1999-08-27									
; PRIOR APPLICATION NUMBER: JP 2000-118776									
; PRIOR FILING DATE: 2000-01-11									
; PRIOR APPLICATION NUMBER: JP 2000-183767									
; PRIOR FILING DATE: 2000-05-02									
; PRIOR APPLICATION NUMBER: JP 2000-241899									
; PRIOR FILING DATE: 2000-06-09									
; PRIOR APPLICATION NUMBER: 60/159,590									
; PRIOR FILING DATE: 1999-10-18									
; PRIOR APPLICATION NUMBER: 60/183,322									
; PRIOR FILING DATE: 2000-02-17									
; NUMBER OF SEQ ID NOS: 19025									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 18965									
; LENGTH: 713									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-917-503-18965									
Query Match 59.6%; Score 2213; DB 35; Length 713;									
Best Local Similarity 64.8%; Pred. No. 5.8e-183;									
Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;									
Qy	1	METGTHRAKRKPGRLGSMFRLPFL-RRSHACSEFPFPPSSRQPNPN-----SALPERRT	54						
Db	1	MEPGTGSRRKRLGPRAGFRFWPPFRRSQAGSSKFPTPLGPENSGNPTLLSSAQPETRV	60						
Qy	55	RYWTKLLSLLALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRESDAPTV	114						
Db	61	SYWTKLLSLLALLPSLFQKLLWSQLSGGLIPTRWLDFAAGVYSALRALGREKPAAPTA	120						
Qy	115	QKSLSYTAAGLFAKTRVVTALARGCTPVAVLVRLVLEVKLKAQERALDSAAPTFLLEQQ	174						
Db	121	QKSLSSQLD-SSDPVTSPLDWLEEGHWHQVSPDILKLEKAGSALDPAQAFLLEQQ	179						
Qy	175	LWGVELLPSSLAQGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQ	234						
Db	180	LWGVELLPSSLAQGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQ	239						
Qy	235	SSAGGQFVGFRTLTPESSCYLSEDCHPQPLRAEMSAATAWRCPPPLSTEGLPPIHHRMR	294						
Db	240	NSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWQGCPLSTEGLPPIHHRMK	299						
Qy	295	WLVPFL-OPNQGQDLPTLDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP----	349						
Db	300	RLEFLQAASKQDLPTPDQDNGYHSLSEEHNLRLMDPKHCRDNPQTFVPAAGDIPGNTQE	359						
Qy	350	PTKKKPELVITQEV-----SQSPQSSSLFCELPVEKECEDHTNATDLSRGESLPVSTRP	404						
Db	360	STEEKIELLTTEVPLALEEESPSCSPSEIPEKEPGEGRISVVDVSYLEGDLPIGARP	419						
Qy	405	VCSNKLIDYILGGAPSDLEASSDSESDGCEEDGDFDSDGSLSESDVDQDSGLHLWN	464						
Db	420	ACSNKLIDYILGGASSDLETSSDPEGEDWDEAEDDGFDSDSLSDSLDQEPGLHLWN	479						
Qy	465	SFHSVDPYKPNQNTATTQTAARIAPRDPDSGTSWSGSCGV-GSCQSGPLPETPDHSSGE	523						
Db	480	SFCSVDPYKPNQNTATTQTAARIIVPEPSDSEKDLGKSDLENSQSGSLPETPEHSSGE	539						
Qy	524	EDDWEPSSADEAENLKLWNSFCHSEDPYNLLNFKAPFPQSGKNWKGQDSKASSEVTVAFS	583						
Db	540	EDDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCDSKTPSESIIVS	599						
Qy	584	GHHTLLSCKAQLLESQEDNCPGCLGEGALAGERTYTHIKRKKVTFLEEVTEYISGDEDRK	643						
Db	600	ECHTLLSCVKQLGSESCEPDSVQRDVLSSGRHTRVKKRKKVTFLEEVTEYISGDEDRK	659						

Qy	644	GPWEFARDGCRFQKRIQETEAIGYCLAPEHREKMFNRLR	684						
Db	660	GPWEFARDGCRFQKRIQETEDAIGYCLTPEHRRMFNRLQ	700						
RESULT 7									
US-09-757-028-2348									
; Sequence 2348, Application US/09757028									
; GENERAL INFORMATION:									
; APPLICANT: Rosen et al.									
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies									
; FILE REFERENCE: PM001									
; CURRENT APPLICATION NUMBER: US/09/757,028									
; CURRENT FILING DATE: 2001-01-09									
; PRIOR APPLICATION NUMBER: 60/179,065									
; PRIOR FILING DATE: 2000-01-31									
; PRIOR APPLICATION NUMBER: 60/180,628									
; PRIOR FILING DATE: 2000-02-04									
; NUMBER OF SEQ ID NOS: 2660									
; SOFTWARE: PatentIn Ver. 2.0									
; SEQ ID NO 2348									
; LENGTH: 720									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-09-757-028-2348									
Query Match 59.5%; Score 2208; DB 22; Length 720;									
Best Local Similarity 64.8%; Pred. No. 1.6e-182;									
Matches 454; Conservative 54; Mismatches 175; Indels 18; Gaps 7;									
Qy	1	METGTHRAKRKPGRLGSMFRLPFL-RRSHACSEFPFPPSSRQPNPN-----SALPERRT	54						
Db	8	MEPGTGSRRKRLGPRAGFRFWPPFRRSQAGSSKFPTPLGPENSGNPTLLSSAQPETRV	67						
Qy	55	RYWTKLLSLLALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASGRESDAPTV	114						
Db	68	SYWTKLLSLLALLPSLFQKLLWSQLSGGLIPTRWLDFAAGVYSALRALGREKPAAPTA	127						
Qy	115	QKSLSYTAAGLFAKTRVVTALARGCTPVAVLVRLVLEVKLKAQERALDSAAPTFLLEQQ	174						
Db	128	QKSLSSQLD-SSDPVTSPLDWLEEGHWHQVSPDILKLEKAGSALDPAQAFLLEQQ	186						
Qy	175	LWGVELLPSSLAQGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQ	234						
Db	187	LWGVELLPSSLAQGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQ	246						
Qy	235	SSAGGQFVGFRTLTPESSCYLSEDCHPQPLRAEMSAATAWRCPPPLSTEGLPPIHHRMR	294						
Db	247	NSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWQGCPLSTEGLPPIHHRMK	306						
Qy	295	WLVPFL-OPNQGQDLPTLDQDNGYHSLSEEHNLRLMDPOHCTDNPAAQVSPAADRP----	349						
Db	307	RLEFLQAASKQDLPTPDQDNGYHSLSEEHNLRLMDPKHCRDNPQTFVPAAGDIPGNTQE	366						
Qy	350	PTKKKPELVITQEV-----SQSPQSSSLFCELPVEKECEDHTNATDLSRGESLPVSTRP	404						
Db	367	STEEKIELLTTEVPLALEEESPSCSPSEIPEKEPGEGRISVVDVSYLEGDLPIGARP	426						
Qy	405	VCSNKLIDYILGGAPSDLEASSDSESDGCEEDGDFDSDGSLSESDVDQDSGLHLWN	464						
Db	427	ACSNKLIDYILGGASSDLETSSDPEGEDWDEAEDDGFDSDSLSDSLDQEPGLHLWN	486						
Qy	465	SFHSVDPYKPNQNTATTQTAARIAPRDPDSGTSWSGSCGV-GSCQSGPLPETPDHSSGE	523						
Db	487	SFCSVDPYKPNQNTATTQTAARIIVPEPSDSEKDLGKSDLENSQSGSLPETPEHSSGE	546						
Qy	524	EDDWEPSSADEAENLKLWNSFCHSEDPYNLLNFKAPFPQSGKNWKGQDSKASSEVTVAFS	583						
Db	547	EDDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCDSKTPSESIIVS	606						
Qy	584	GHHTLLSCKAQLLESQEDNCPGCLGEGALAGERTYTHIKRKKVTFLEEVTEYISGDEDRK	643						

Qy	417	GA	P	S	D	L	E	A	S	S	D	S	E	S	D	E	D	G	P	D	S	G	S	L	S	S	D	V	E	Q	S	E	G	L	H	L	W	N	S	F	H	S	V	D	P	Y	K	P	K	O	N	476						
Db	426	GA	S	S	D	L	E	T	S	S	D	P	E	G	E	D	D	E	A	D	G	F	D	S	L	S	D	S	L	E	Q	P	E	G	L	H	L	W	N	S	F	C	S	V	D	P	Y	N	P	O	N	485						
Qy	477	FT	A	T	I	Q	T	A	A	R	I	A	P	R	P	S	D	S	G	T	S	W	S	G	C	G	V	-	G	S	C	Q	E	G	P	L	P	E	T	P	H	S	G	S	E	E	D	M	P	S	A	D	E	A	535			
Db	486	FT	A	T	I	Q	T	A	A	R	I	V	P	E	P	S	D	S	E	K	D	L	S	K	S	D	L	E	N	S	Q	S	G	L	P	E	T	P	H	S	G	S	E	E	D	M	E	S	S	A	D	E	A	545				
Qy	536	N	L	K	L	W	N	S	F	C	H	S	D	P	N	L	N	L	N	F	K	A	P	O	P	S	K	N	K	W	G	R	Q	D	S	K	A	S	S	E	V	T	A	F	S	G	H	T	L	L	S	C	K	A	O	L	595	
Db	546	S	L	K	L	W	N	S	F	C	N	S	D	P	N	L	N	L	N	F	K	A	P	O	T	S	E	N	E	K	G	R	D	S	K	T	P	S	E	S	I	V	A	I	S	E	C	H	T	L	L	S	C	K	V	O	L	605
Qy	596	L	S	E	Q	E	N	C	P	E	C	G	L	B	A	L	A	G	E	R	Y	T	H	I	K	R	K	V	T	F	L	E	B	V	T	E	Y	I	I	S	G	D	E	R	K	G	P	W	E	F	A	R	D	G	C	R	655	
Db	606	L	G	S	E	B	E	C	P	D	S	V	Q	R	D	L	S	G	G	R	H	T	H	K	R	K	V	T	F	L	E	B	V	T	E	Y	I	I	S	G	D	E	R	K	G	P	W	E	F	A	R	D	G	C	R	665		
Qy	656	F	O	K	R	I	O	E	T	E	V	A	I	C	Y	C	L	A	P	E	H	R	E	K	M	N	L	R	684																													
Db	666	F	O	K	R	I	O	E	T	E	D	A	I	G	Y	C	L	T	E	H	R	E	R	M	N	L	O	694																														

```

RESULT 10
US-09-925-299-941
; Sequence 941, Application US/09925299
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

```

Qy	247	TLTPTESC	YLS	EDG	CH	PQ	LRA	EM	AT	AW	RC	CP	PL	ST	E	G	L	P	E	I	H	H	R	M	R	W	L	V	L	-	Q	N	Q	G	305	
Db	246	TLTPTESC	LR	DH	CH	PQ	LRA	EM	AT	AW	RC	CP	PL	ST	E	G	L	P	E	I	H	H	R	M	R	W	L	V	L	-	Q	N	Q	G	305	
Qy	306	DLPTL	D	D	O	O	N	G	V	H	S	L	E	E	E	H	N	L	R	M	D	P	O	H	C	T	D	N	P	A	Q	A	V	S	361	
Db	306	DXPTP	D	O	O	N	G	V	H	S	L	E	E	E	H	N	L	R	M	D	P	O	H	C	T	D	N	P	A	Q	A	V	S	365		
Qy	362	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	416			
Db	366	VPLA	E	E	E	S	P	E	G	C	P	S	S	E	I	P	M	E	K	P	E	G	R	I	S	V	D	Y	S	L	E	G	D	L	425	
Qy	417	GAP	S	D	L	R	A	S	S	D	S	E	S	E	D	D	G	F	D	S	G	L	S	E	S	D	V	E	O	D	S	E	G	L	H	476
Db	426	GASS	D	L	E	T	S	S	P	E	G	E	D	M	D	E	E	A	E	D	G	F	D	S	S	L	S	D	L	E	Q	D	P	E	485	
Qy	477	FTAT	I	O	T	A	A	R	I	A	P	R	D	S	D	G	T	S	W	S	G	C	V	-	G	S	C	O	E	G	P	L	P	E	535	
Db	486	FTAT	I	O	T	A	A	R	I	V	P	E	P	S	D	S	E	K	D	L	S	G	S	O	L	E	N	S	S	O	G	S	L	P	545	
Qy	536	NK	L	W	N	S	F	C	H	S	E	D	P	N	L	I	N	F	K	A	P	O	P	S	K	N	K	G	R	O	D	S	K	A	S	595
Db	546	SK	L	W	N	S	F	C	N	S	D	D	P	N	L	I	N	F	K	A	P	O	T	S	G	E	N	E	K	G	R	D	S	T	605	
Qy	596	LES	O	E	D	N	C	P	G	G	L	E	A	L	A	G	E	R	Y	T	H	I	K	R	K	V	T	F	L	E	E	V	T	Y	I	655
Db	606	LGS	O	E	S	E	C	P	D	S	V	Q	R	D	L	S	G	R	H	T	H	V	K	R	K	V	T	F	L	E	E	V	T	Y	I	665
Qy	656	FO	K	R	I	O	E	T	A	I	G	Y	C	L	A	F	E	H	R	E	K	M	P	N	R	L	R	684								
Db	666	FO	K	R	I	O	E	T	A	I	G	Y	C	L	T	F	E	H	R	E	R	M	F	N	R	L	Q	694								

```

RESULT 11
PCT-US00-35017A-963
; Sequence 963, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-35017A-963

```

RESULT 12
US-10-296-115-963
; Sequence 963, Application US/10296115

```
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-963

Query Match      14.7%; Score 547; DB 28; Length 153;
Best Local Similarity 76.6%; Pred. No. 1.4e-38;
Matches 105; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

QY 548 DPYNLLNFKAPQPSGKNWKGRODSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPCG 607
DB 4 DPYNPLNFKAPQTSGENEKGCRDSTPSEIVATSECHTLLSCKVQLLGSQSECPDSV 63

QY 608 LGEALAGERYTHIKRKKVTFLEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETVAI 667
DB 64 QRDVLSSGRHVTAKKVTFLFEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETDAI 123

QY 668 GYCLAFHREKMFNRLR 684
DB 124 GYCLTFEHRERMFNRLQ 140

RESULT 13
PCT-US01-14827-15743
; Sequence 15743, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 15743
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-14827-15743

Query Match      9.6%; Score 357; DB 1; Length 101;
Best Local Similarity 75.0%; Pred. No. 2.7e-22;
Matches 66; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 597 ESQEDNCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYYISGDEDRKGPWFEPFARDGCRF 656
DB 1 KSQSECPDSVQORDVLSSGRHVTAKKVTFLFEEVTEYYISGDEDRKGPWFEPFARDGCSV 60

QY 657 QKRIQETVAIGYCLAFHREKMFNRLR 684
DB 61 QKRIQETDAIGYCLTFEHRERIGNRLQ 88

RESULT 14
US-60-196-718-7356
; Sequence 7356, Application US/60196718
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
```

```
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000456
; CURRENT APPLICATION NUMBER: US/60/196,718
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7356
; LENGTH: 67
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-718-7356

Query Match      8.3%; Score 307; DB 37; Length 67;
Best Local Similarity 93.3%; Pred. No. 3.2e-18;
Matches 56; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 625 VTFLFEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETVAIGYCLAFHREKMFNRLR 684
DB 1 VTFLFEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLQ 60

RESULT 15
US-60-177-571-4491
; Sequence 4491, Application US/60177571
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000201
; CURRENT APPLICATION NUMBER: US/60/177,571
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 5082
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4491
; LENGTH: 73
; TYPE: PRT
; ORGANISM: HUMAN
US-60-177-571-4491

Query Match      8.3%; Score 307; DB 37; Length 73;
Best Local Similarity 93.3%; Pred. No. 3.7e-18;
Matches 56; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 625 VTFLFEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETVAIGYCLAFHREKMFNRLR 684
DB 1 VTFLFEEVTEYYISGDEDRKGPWFEPFARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLQ 60

Search completed: September 16, 2005, 10:53:31
Job time : 182.581 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:42:43 ; Search time 84.5911 Seconds
(without alignments)
1348.829 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRARKRPGRLGSWF.....MFNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 760079 seqs, 163465437 residues

Total number of hits satisfying chosen parameters: 760079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Pending_Patents_AA_New: *
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	189.5	5.1	674	7	US-11-040-219-2
2	175.5	4.7	605	7	US-11-040-219-5
3	151	4.1	3880	1	PCT-US04-37982-2437
4	151	4.1	3880	1	PCT-US04-37982-2439
5	148.5	4.0	676	7	US-11-097-143-13533
6	140.5	3.8	916	6	US-10-899-422-13
7	140.5	3.8	1059	6	US-10-899-422-11
8	139.5	3.8	883	1	PCT-US05-01184A-6
9	138	3.7	913	1	PCT-US04-37982-188
10	138	3.7	913	1	PCT-US04-37982-202
11	138	3.7	1084	1	PCT-US04-37982-192
12	138	3.7	1572	6	US-10-990-328A-11742
13	136.5	3.7	883	1	PCT-US05-01184A-3
14	136.5	3.7	3926	8	US-60-675-841-94
15	136.5	3.7	3926	8	US-60-710-726-316
16	132.5	3.6	1065	6	US-10-940-774A-11618
17	132	3.6	1715	7	US-11-097-143-1059
18	127.5	3.4	3913	6	US-10-940-774A-10933
19	127.5	3.4	4232	8	US-60-651-506-210
20	127.5	3.4	4377	1	PCT-US04-42360-2065
21	127.5	3.4	4377	8	US-60-651-506-211
22	127.5	3.4	4377	8	US-60-651-506-212
23	127	3.4	1028	6	US-10-531-164-756
24	127	3.4	1503	1	PCT-US05-18850-797
25	124.5	3.4	1354	7	US-11-097-143-13869

26	124.5	3.4	1709	1	PCT-US05-21650-40
27	124	3.3	640	6	US-10-450-763-41872
28	124	3.3	1795	6	US-10-450-763-51377
29	123.5	3.3	4274	6	US-10-450-763-31331
30	123.5	3.3	4377	6	US-10-940-774A-6978
31	123.5	3.3	4377	8	US-60-651-506-209
32	123.5	3.3	4386	6	US-10-450-763-37734
33	123.5	3.3	4397	6	US-10-450-763-52303
34	123	3.3	1142	7	US-11-044-051-73
35	123	3.3	1155	6	US-10-450-763-54525
36	123	3.3	1208	6	US-10-450-763-54526
37	122.5	3.3	808	7	US-11-072-512-3401
38	122.5	3.3	961	1	PCT-US03-10870-1618
39	122	3.3	1342	1	PCT-US03-10870-2821
40	121.5	3.3	530	7	US-11-043-591-168
41	121.5	3.3	639	7	US-11-043-591-169
42	121.5	3.3	671	7	US-11-043-591-167
43	120.5	3.2	306	8	US-60-707-841-176
44	120	3.2	1261	6	US-10-940-774A-9651
45	120	3.2	1393	6	US-10-994-106-23
46	120	3.2	1393	6	US-10-994-106-47
47	120	3.2	1393	6	US-10-994-106-58
48	120	3.2	1393	6	US-10-994-106-345
49	120	3.2	1684	1	PCT-US05-00638-84
50	120	3.2	1686	8	US-60-680-002-257
51	120	3.2	1782	1	PCT-US03-10870-2097
52	120	3.2	1784	8	US-60-680-002-256
53	119.5	3.2	1316	6	US-10-450-763-53356
54	119.5	3.2	1321	6	US-10-450-763-55775
55	119	3.2	535	8	US-60-655-875-120587
56	119	3.2	665	7	US-11-043-591-166
57	119	3.2	1040	6	US-10-450-763-51333
58	118.5	3.2	3896	1	PCT-US04-37982-2434
59	118	3.2	1168	6	US-10-990-328A-13049
60	118	3.2	1239	6	US-10-990-328A-13050
61	118	3.2	1239	6	US-10-990-328A-13051
62	118	3.2	1781	8	US-60-680-002-258
63	117.5	3.2	992	7	US-11-090-997-56
64	117.5	3.2	1188	7	US-11-090-997-52
65	117.5	3.2	1413	1	PCT-US04-31416-52
66	117.5	3.2	1423	6	US-10-990-328A-12800
67	117.5	3.2	1698	1	PCT-US04-31416-54
68	117.5	3.2	1709	6	US-10-479-874A-21
69	117.5	3.2	1710	6	US-10-990-328A-12801
70	117.5	3.2	1712	6	US-10-940-774A-9450
71	117.5	3.2	1970	7	US-11-097-143-21373
72	117	3.2	422	7	US-11-072-512-2882
73	117	3.2	780	6	US-10-489-448-3405
74	117	3.2	819	1	PCT-US03-10870-1793
75	117	3.2	1051	7	US-11-097-143-7263
76	117	3.2	1411	8	US-60-710-726-1468
77	116.5	3.1	535	6	US-10-450-763-36563
78	116.5	3.1	535	6	US-10-450-763-44703
79	116.5	3.1	535	6	US-10-450-763-57981
80	116.5	3.1	1454	1	PCT-US03-10870-1328
81	116	3.1	742	6	US-10-538-002-222
82	116	3.1	817	6	US-10-285-830A-8
83	116	3.1	817	7	US-11-041-102-8
84	116	3.1	2312	7	US-11-126-313-34
85	115	3.1	805	6	US-10-940-774A-6713
86	115	3.1	805	7	US-11-037-713-24
87	115	3.1	1468	7	US-11-097-143-15765
88	114.5	3.1	654	6	US-10-781-060-16
89	114.5	3.1	797	6	US-10-204-639-49
90	114.5	3.1	797	6	US-10-204-639A-49
91	114.5	3.1	901	6	US-10-330-773A-870
92	114.5	3.1	1163	7	US-11-097-143-1281
93	114.5	3.1	1210	7	US-11-051-454-264
94	114.5	3.1	1210	8	US-60-710-726-530
95	114	3.1	503	6	US-10-703-032-123990
96	114	3.1	605	5	US-09-833-245B-2237
97	114	3.1	605	6	US-10-934-272-2
98	114	3.1	734	7	US-11-097-143-33975

Sequence 40, Appl
Sequence 41872, A
Sequence 51377, A
Sequence 31331, A
Sequence 6978, Ap
Sequence 209, App
Sequence 37734, A
Sequence 52303, A
Sequence 73, Appl
Sequence 54525, A
Sequence 54526, A
Sequence 3401, Ap
Sequence 1618, Ap
Sequence 2821, Ap
Sequence 168, App
Sequence 169, App
Sequence 167, App
Sequence 176, App
Sequence 9651, Ap
Sequence 23, Appl
Sequence 47, Appl
Sequence 58, Appl
Sequence 345, App
Sequence 84, Appl
Sequence 257, App
Sequence 2097, Ap
Sequence 256, App
Sequence 53356, A
Sequence 55775, A
Sequence 120587, A
Sequence 166, App
Sequence 51233, A
Sequence 2434, Ap
Sequence 13049, A
Sequence 13050, A
Sequence 13051, A
Sequence 13051, A
Sequence 358, App
Sequence 58, Appl
Sequence 56, Appl
Sequence 52, Appl
Sequence 12800, A
Sequence 54, Appl
Sequence 21, Appl
Sequence 12801, A
Sequence 9450, Ap
Sequence 21273, A
Sequence 2882, Ap
Sequence 3405, Ap
Sequence 1793, Ap
Sequence 7263, Ap
Sequence 1468, Ap
Sequence 36563, A
Sequence 44703, A
Sequence 57981, A
Sequence 1328, Ap
Sequence 222, App
Sequence 8, Appl
Sequence 34, Appl
Sequence 6713, Ap
Sequence 24, Appl
Sequence 15765, A
Sequence 16, Appl
Sequence 49, Appl
Sequence 49, Appl
Sequence 870, App
Sequence 1281, Ap
Sequence 264, App
Sequence 530, App
Sequence 123990, A
Sequence 2237, Ap
Sequence 2, Appl
Sequence 33975, A

```
99 114 3.1 802 7 US-11-090-997-768 Sequence 768, App
100 114 3.1 1067 7 US-11-097-143-13260 Sequence 13260, A

ALIGNMENTS

RESULT 1
US-11-040-219-2
; Sequence 2, Application US/11040219
; GENERAL INFORMATION:
; APPLICANT: Paul Fisher
; TITLE OF INVENTION: Chimeric tumor suppressor gene and
; FILE REFERENCE: 36351 (070050.2617)
; CURRENT APPLICATION NUMBER: US/11/040,219
; CURRENT FILING DATE: 2005-01-21
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; OTHER INFORMATION: human GADD34 amino acid sequence
US-11-040-219-2

Query Match 5.1%; Score 189.5; DB 7; Length 674;
Best Local Similarity 21.5%; Pred. No. 0.0012;
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;

QY 60 LLSQALLPSLFQKLLWSQLSGGLIPTR-WLDPFAASYSALRASRGRESDAPVQKSL 118
DB 21 LLSPPVGLLSR-----AWSRLR-GLGLEPFWLVAVKGAAL----- 55

QY 119 SYTAAGLFAKTRVVTALARGGTPVAVLRLRLEVKLKAQBRALDSAP-----T 168
DB 56 --VEAGLGEAR-----IPLA--IPIHTWGRPREPEAEEDSGGPGEDRETGLTKT 100

QY 169 FLLSQQLMGVBLLPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSVLLNPSYL----- 223
DB 101 SSSLPEAWG--LLDD--DDGMGEREATSVPRGQSQPADGQ-----RAPLSFSLIRTLQ 152

QY 224 --DVLPLQLGLRCQSSAGGQFVGPRTLT-----PESCYLSDDGCHPOPLRAEMSATA--- 273
DB 153 GSDKNP--GEKAEHEGVAEEGVNFKVYPPSHRECCPAVEEEDDEEAKVKEAHTSTSA 210

QY 274 -----WRRCP-----PLSTEGLPETIHHRRMRWLVLQPNQGDLPPLDQNGYHSLE 320
DB 211 LSPGSKPTWVSCPGEEENQATEDKRTSRKGAARTSVSPRSSGSDPRSWEYRSGEASEE 270

QY 321 EEHLLRMDPOHCTDNPAQAVSPAADRPETPEKPELVIQEVSQSPQSSSLFCBLPVEKE 380
DB 271 KE-----EKAHETKGEEA-----PGQSSAP-----AQRPLKSWWCQPSDEEE 311

QY 381 CEEDHTNATLSDRGESLPVSTRPVCSNKLIDYILGGAPSDLEASSDSESDWGEPEDD 440
DB 312 GEVALGHAEKDGEACPCIPPSAFLKAWVYFG-----EDTEEEDEE-----DESDS 364

QY 441 GFSDGSLSESDEVDQSEGLHL--WNSFHSVDYPKQNFNTATIQ-----TAARIAPRDS 493
DB 365 GSDEEGEAEASSSTPATGVFLKSW-----VYQPGEDTEEEDESDTGSADEREAE 417

QY 494 DSGTSSWSCGVCQCEGFLPETPDH-----SSGEEDDWEPADAEANLKLWNSFCHSE 547
DB 418 TSASTPASAFLKAWVYRPGEDTEEEDEEDVDSEKDDSEALGEAES----- 466

QY 548 DPYNLLNFKAPFPQSG-----KNWKGRDQSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603
DB 467 DPH-----PSHPDQRAHFRGMYRPGKETEE-----EAEED-- 498

QY 604 PGCGLGEA-----LAGER-----YTHIKRK 623
```

```
DB 499 ----WGEAECPPFRVAIVPGCEKPPPPWAPPRLPLRLQRLKRPEPTHDPDPETPLKAR 554
QY 624 KVTFLLEVITYIS-----GDEDRKGPWFBSFARDGCRFKRIQETEVAGYCLAFAHREK 678
DB 555 KVFSEKVTYVHFLAVWAGPAQAARQGPWEQLARDRSFARRITQAOBELSPCLTAAARAR 614
QY 679 MFNRLR 684
DB 615 AWAELR 620

RESULT 2
US-11-040-219-5
; Sequence 5, Application US/11040219
; GENERAL INFORMATION:
; APPLICANT: Paul Fisher
; TITLE OF INVENTION: Chimeric tumor suppressor gene and
; FILE REFERENCE: 36351 (070050.2617)
; CURRENT APPLICATION NUMBER: US/11/040,219
; CURRENT FILING DATE: 2005-01-21
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: rat/human chimera (artificial) amino acid sequence
; OTHER INFORMATION: of rPEG-3 1-347 + hGADD34 422-674
US-11-040-219-5

Query Match 4.7%; Score 175.5; DB 7; Length 605;
Best Local Similarity 21.6%; Pred. No. 0.0077;
Matches 131; Conservative 61; Mismatches 203; Indels 211; Gaps 31;

QY 216 YLLNPSYLDLPQLGLR-----COSSAGGQFVGFRITLTPESCYLSDEGCHPQP 264
DB 20 YLLSP-LMGFLSRAWSRLRGPEVSEAWLAETVAGANQIEADALLTTPP--VSENHL---P 73

QY 265 LR-----AEMSATAWRC-----PPSTEGLPETIHHRRMRWLVLQPNQGDLPPL 310
DB 74 LRETEGNTPEWSKAAQRLCLDVEAQSPPKTWGLSDIDEHNGK-----PGQD--- 121

QY 311 DDNGYHSLEEH--NLLRMDPOHC--TDNPAQAV-----SPAADR 347
DB 122 ----GLRQEVEHTAGLPTLOPLHLQADKKVGEVVARBEGVSELAYPTSHWEGGPADE 177

QY 348 PETEKPELVIQEVSQSP---QGSSSLFCBLPVEKECED-----HTNATDLS 392
DB 178 EDTETVKAHQAAASAIAPIGVKPSVYVCPGEAEHRAATEKEGTDNKAEPSSGSHSRVWEYH 237

QY 393 DR-----GSSLPVSTRP-----VCSNKLIDYILGGAPSDLEASSDSESDW-----GEPE 438
DB 238 TRERPKEGETPEKPEHRAQSHPCNAEAE---EGGPETSVCSGSAFLKAWVYRPGEDTE 294

QY 439 DDGFDSDGSLSESDEVDQSEGLHLWNSFHSVDYPKQNFNT-----ATIQAARIAP 489
DB 295 EEE--DSDLSAEDEAHTCTTPT--SAFLKAWVYRPGEDTEEEEDDGDWDSAEEDALTSAS 352

QY 490 RDPDSDG--TSWSGSCGVCQCEGFLPETPD--HSSGEEDDWEPADAEANLKLWNSFCHSE 547
DB 353 TPASAFKAWVYRPGEDTEEE---EDVDSEKDDSEALGEAES----- 397

QY 548 DPYNLLNFKAPFPQSG-----KNWKGRDQSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603
DB 398 DPH-----PSHPDQRAHFRGMYRPGKETEE-----EAEED-- 429

QY 604 PGCGLGEA-----LAGER-----YTHIKRK 623
DB 430 ----WGEAECPPFRVAIVPGCEKPPPPWAPPRLPLRLQRLKRPEPTHDPDPETPLKAR 485
QY 624 KVTFLLEVITYIS-----GDEDRKGPWFBSFARDGCRFKRIQETEVAGYCLAFAHREK 678
```


[illegible][illegible]

Query Match.	3.8%;	Score 140.5;	DB 6;	Length 916;
Best Local Similarity	19.4%;	Pred. No. 2;		
Matches 166;	Conservative 79;	Mismatches 250;	Indels 361;	Gaps 36;
QY	5	THRARKEPGR-----LGSWRF-----LPLRRSHACSEPPPP	38	
Db	130	THGVQARPEQRHSHLVCDSDGLLLGWERYETVAKRGGOHTLPALSRATPDNPVII	189	
QY	39	SSRQNPNSALPERRTRYTKLLSQLLALLPSLFQKILLMSQLSGGJLPTRWLDFAA SYS	98	
Db	190	LAPLNPQSQ-----LLRPVYQE-----DIIP-----	210	
QY	99	ALRASRGRESDPATVOKSLSYTAAGLPFAKTRVVVSTLALARG-----	140	
Db	211	-----EBSGSPS-----KGKSYT-----GLGKSKRLMKTQTMKGHGYNQPCVVRPHATHSS	258	
QY	141	-GTPVAVILRLVKAKAQRALDSAPTFLEQOLAGVELLPSSLOAGLVSHREL----	195	
Db	259	YGYTVTAPKVLVPFVTFQPDLCNPARTLLLSSELLLYEGRNKAEEVTLFAYSDLLFT	318	
QY	196	DSSSGSPLSYQSLGNFKVSYLLNPSYLDLPLQGLRCQSSAGGQFVGRFTLTPESCYL	255	
Db	319	KEDEPGRCDV-----LRNPLYQSV-----KLQEGSSDLKFCVL-----YL	355	
QY	256	SEBGCHPOPIRAEMSAATAWRC-----PPLSTEGLEPIHHRMRWL----	296	
Db	356	AEKAECLFTLEAHSQEOKRVCWCLSENIAKQQOALAAASPPDKMFTEADEKREMALEEG	415	
QY	297	-----VFLQPNQODL-----PTLDQDNGYHSLREEHLLRMD-----	329	
Db	416	KPGAEDSPSKSPFS PQEULP PQODLPNKDPSGQEPAPSOEPLSSKDSATSGSPGP	475	
QY	330	-----PQHCTDNPAQAVSPAADRP-----EPTEKKPELVIQE-----VSQSPQGSLSLFC	374	

Db 476 DAPPSKDVPPCQBPAPPAQDLSPCQDLPAQOEPLPHQDPLLTOKLPAIQESPTRDLPPCQD 535
Qy 375 -----LPVEKECEEDHTNATLSDRGESLPVSTRPV-----C 406
Db 536 LPPSQVSLPA-KALTEDTMSSGDLAATGPPAPPAFVPIPEVRLDSTYSQKAGAEQGC 594
Qy 407 SNKLIDYILGAPSDLEASDSSESDEWGPEDDGFDS--GSLSES-----DVEQDSE 458
Db 595 S-----GDEDAEAEVEBEGEGEDEDSDTSDNYGERSEAKRSMIETGQAE 645
Qy 459 G-----LHLWNSF-----HSDVDYKQNFATTIQTAAARIAPDRP-----SDSGTWSGCGVGC 508
Db 646 GGLSLRVQNSLRRRTHS-----EGSLLQEPGCPFASDTTLHCSDGEGAAT 692
Qy 509 QEGPLPET-----PDHSS-----GEEDD----- 526
Db 693 WGMPSPTLKKELGRNGSMHLSLFTGHRKMSGADTVGDDEASRKRKSKNRAKDMKN 752
Qy 527 -----WEPASDEA-----ENLKLWNSFCHS 546
Db 753 KLGIFRRRNEBSPAGPAGKADMMKSFRTSEALKWGESLEKLLVHKDGLAVFQAFRLT 812
Qy 547 EDPVNLNFKAPPQPSCKNW-----KGRDQSKASSEVTVAFGHHTLLSCKAQLLES- 598
Db 813 EFSSENLFF-----WLACEDFKVKVQSKQWASKAKKIFABYIAIQACKEVNLDYS 862
Qy 599 -----QEDNCP-----GC 606
Db 863 TREHTKDNLSQSVTRGC 878

RESULT 7
US-10-899-422-11
; Sequence 11, Application US/10899422
; GENERAL INFORMATION:
; APPLICANT: Lorientis limited
; TITLE OF INVENTION: Medical Treatment
; FILE REFERENCE: P020742US CLM
; CURRENT APPLICATION NUMBER: US/10/899,422
; CURRENT FILING DATE: 2004-07-26
; PRIOR APPLICATION NUMBER: GB0201674.9
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1059
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-899-422-11

Query Match 3.8%; Score 140.5; DB 6; Length 1059;
Best Local Similarity 19.2%; Pred. No. 2.4;
Matches 165; Conservative 82; Mismatches 243; Indels 371; Gaps 37;
Qy 5 THRARKPGPR-----LGSWFR-----LPFLRRSHACSSSEPPPP 38
Db 273 THGVQARPEQRHCHLVCDSDGELLGWERYTEVAKRGQHTLPALSRATAPDPNVII 332
Qy 39 SSRONPGNSALPERRTRYWTKLSQLLALPSLFQKLLWSQLSGGLIPTRWLDFASYS 98
Db 333 LAPLNPQS-----LRPVYQE-----DTIP----- 353
Qy 99 ALRASRGRESDAPTOKLSYTAAGLFAKTRVYSTLALARG----- 140
Db 354 -----EEGSGFS--KGKSYT--GLGKSRIMKTVTQMKGHYQNCVPRPHATHSS 401
Qy 141 -GTPVAVLVLRLEVKLKAQERALDSAAFTFLLEQLQMGVELLPSSLOAGLVSHREL----- 195
Db 402 YGTVVTLAPKLVLPVFPVQPLDLCNPARTLLSBEULLYERKNAEAVTLFPAYSDDLFT 461
Qy 196 DSSSGSPLSVQSLGNFKRVSVLLNPSYLDYLPQLGLRCQSSAGGQGVGFRTLTPECYL 255
Db 462 KEDEBGRCDV-----LRNPLYLSV-----KLEGGSSSDDLKFCVL-----YL 498

Qy 256 SEDGCHPOPLRAEMSAATAWRR-----PPLSTEGLPETIHHRRMRWLVL 299
Db 499 AEKAECLFTLEAHSQKRVKWCLSSENTAKOQOLAASPPDSKKLHTDADHKE-----M 553
Qy 300 QPNQ-----GQDL-PTLDODNGYHSLSEHNLLRMD----- 329
Db 554 APDKANKGAEDSPSPKEPSPQELPPGQDLPPNKKDPSGQEPAPQOEPLSSKDSATSEG 613
Qy 330 -----PQCHTNPQAQVSPAADRP-----BPTKKKPELVIQE-----VSQSPQSS 370
Db 614 SPPGPDAPSKDVPQCEPPPAQDLSPCQDLPAQOEPLPHQDPLLTOKLPAIQESPTRDL 673
Qy 371 LFCE-----LPVKECEEDHTNATLSDRGESLPVSTRPV----- 405
Db 674 PCQDLPPSQVSLPA-KALTEDTMSSGDLAATGPPAPPAFVPIPEVRLDSTYSQKAG 732
Qy 406 -----CSNKLDIYLGCAPSDLEASDSSESDEWGPEDDGFDS--GSLSES-----DV 453
Db 733 AEGGCS-----GDEDAEAEVEBEGEGEDEDSDTSDNYGERSEAKRSMIET 783
Qy 454 EQDSEG-----LHLWNSF-----HSDVDYKQNFATTIQTAAARIAPDRP-----SDSGTWSGSC 503
Db 784 GQAEGLSLRVQNSLRRRTHS-----EGSLLQEPGCPFASDTTLHCSDGE 830
Qy 504 GVSGCQEGPLPET-----PDHSS-----GEEDD----- 526
Db 831 GAASTWGMPSPTLKKELGRNGSMHLSLFTGHRKMSGADTVGDDEASRKRKSKNRA 890
Qy 527 -----WEPASDEA-----ENLKLWNS 541
Db 891 KOMNKLGIFFRRRNEBSPAGPAGKADMMKSFRTSEALKWGESLEKLLVHKDGLAVPQ 950
Qy 542 SFCHSEDPVNLNFKAPPQPSCKNW-----KGRDQSKASSEVTVAFGHHTLLSCKAQ 594
Db 951 AFLRTFSENLFF-----WLACEDFKVKVQSKQWASKAKKIFABYIAIQACKEV 1000
Qy 595 LLES-----QEDNCP-----GC 606
Db 1001 NLDSYTREHTKDNLSQSVTRGC 1021

RESULT 8
PCT-US05-01184A-6
; Sequence 6, Application PC/TUS0501184A
; GENERAL INFORMATION:
; APPLICANT: Yale University
; APPLICANT: Hockfield, Susan
; APPLICANT: Matthews, Russell T.
; APPLICANT: Viapiano, Mariano S.
; TITLE OF INVENTION: PRIMARY CENTRAL NERVOUS SYSTEM TUMOR SPECIFIC BEHAV ISOFORMS
; FILE REFERENCE: 044574-5126-WO
; CURRENT APPLICATION NUMBER: PCT/US05/01184A
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: US 60/536,594
; PRIOR FILING DATE: 2004-01-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Rattus rattus
; PCT-US05-01184A-6

Query Match 3.8%; Score 139.5; DB 1; Length 883;
Best Local Similarity 22.3%; Pred. No. 2.2;
Matches 154; Conservative 66; Mismatches 217; Indels 253; Gaps 39;
Qy 140 GGTPEVAVLV-----LRLEVKLKAQERALDSAAFTFLLEQLQMGVELLP--SSLOAGLVSHRE 194
Db 83 GDEVEVLVARGLRVKVNEAYRFRVALPAYPASLTDVSLVLSLRPNDSGVYRCEVQHGI 142
Qy 195 LDSSSGSPLSVQSL-----GNFKVSVSVLLNPSYLDYLPQLGLRCQ 234

```
Db 143 DDSSDAVEKVKGVFLYREGSARYAFSAGACARIGARIATPEQL-VAAYLGGVEQ 201
Qy 235 SSAG--GGQVCFRFLTP-ESCYLSEGG-----CHPQLRAEMSA- 271
Db 202 CDAGWLSQDTVRYPIQNPACYGMGYPGVNRYGVVGGDDLYVDVYCYAEDLNGELFLG 261
Qy 272 -----TAWR---RCP-----PLSTE----- 283
Db 262 APPGKLWEEARDYCLERGAQIASTGQLYAAWNGGLDRCSFGWLADGVSRYPIITPSQRC 321
Qy 284 --GLPEIHRMRMLVFLQPNQGDLPDLDQNGVHSLBEEHNLRLMDPQCTONPA--- 338
Db 322 GGGLPGVK-----TLPLFPNQ-----TGPFSKQNRNV-----YCFRDSAHPS 359
Qy 339 ---QAVSPAADRPEP---TEKKPELVI-QEVSQSPQSGSLFCELPVEKECEEDHTNATD 390
Db 360 AFSEASSPASDGLAIVTVTEKLELQPBQAVSESRGAIY-SIPI-----TE 407
Qy 391 LSDRGESLPVSTRPVCSNKLIDYILGGAP-SDLEASSDS-----ESDMDGEEPEDDGFD 443
Db 408 DGGGSSSTPEDP-----AEAPRTLESETQSVAPPTCSSEEGEALBEEERF 454
Qy 444 SDGSLSESDVQDSGLHLW-----NSFHSVDPYKPNFTATIQTAARIAPRD 491
Db 455 KDTETPKKEQKQ--ENLWVWPTLSSPLTGLTEHLSLSQVSP- AQAVLQLGASPPRP 511
Qy 492 PSDGTSWSGCGVSCGEGPLPETPDHS---SCEDDWEPSEADEAENLKLWNSFCHSED 548
Db 512 PRVHGPP---AETQPPREGSLTSTPDGAREVAGETGSPFELSGVPRSEEGSS--SLED 566
Qy 549 PYNLINFKAQTPQSG-----KNWKR-----QDSKASSEVTAVFSGH 585
Db 567 GPSLL--PATWAPVGTRELTPESEKSGRTVLTGTSVQAQVPLPTDSASRGVAVAPSSG 624
Qy 586 HTLIS-CK--AQLLESQED-NC---PG-----CGLG-----EALAGERYTHIKKKV 625
Db 625 DCIPSPCHNGTCLKEEGFRCLCLPGYGGDLCDVGLHFCSPGWEAFQACYKHFSTR-- 683
Qy 626 TFLBEVTEYIIGDEDRKGPWEFARDGCR 655
Db 684 -----SWE-EAESQCR 693

RESULT 9
PCT-US04-37982-188
; Sequence 188, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; PRIOR FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-188

Query Match 3.7%; Score 138; DB 1; Length 913;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRAKRKPGPR-----LGSWFR-----LPFLRRSHACSSERPPP 38
Db 128 THGVOARPEQRHCHLVCDSSGLLGGWERYTEVAKRGQHTLPALSRATPTDPNVII 187
```

```
Qy 39 SSRONPGNSALPERRTRYTKLISQLLALLPSLFQKLLWSQLSGGILPTRWLDFAASYS 98
Db 188 LAPLPSQS-----LLRPVQE-----DIP----- 208
Qy 99 ALRASRGRESDDPTVOKSLSYTAAGLFAKTRVSVSTLALARG----- 140
Db 209 -----EESGSPS--KGKSYT--GLGKESRLMKTVQTMKGHGNYQCPVVRPHATHSS 256
Qy 141 -GTPVAVLVLRLEVKAKQERALDSAAPTYLLEQOLMGVELLPSSLQAGLVSHREL--- 195
Db 257 YGTVVTLAPKVLVPPVFQPLDLCNPARTLLSEELLLYEGRNKAABVTFLPAYSDLLFT 316
Qy 196 DSSSGSLSVQSLGNFKVSVYLLNPSYLDLPQLGLFCQSSAGGQGVGFRTITPESCYL 255
Db 317 KEDEPGRCDV-----LRNPLYQSV-----KLQEGSSEDLKFCVL-----YL 353
Qy 256 SEDG-----CHPQPLR-----AEMSATAWRCPPPLSTEGPLPEIHRMRWL----- 296
Db 354 AEKAECLFTLEAHSQEQKRVWCWCLSENIAKQOQLADPDKMPFETADEKREMALEBKG 413
Qy 297 -----VFLQPNQGDLP-----PTLDQDNGVHSLBEEHNLRLMD----- 329
Db 414 PGABDSPPSKEPSPGQELPPGQDLPPNKSQSPQEPAPSOBPLSSKDSATSESGFPQDPA 473
Qy 330 -----POHCTDPAQAVSPAADRP---BPTKKPELVIQE---VSQSPQSGSLFCE--- 374
Db 474 PPSKDVPPCQEPAPQDLSPCQDLPAQGEPLPHQDPLLTOKDLPAIQESPTRDLPPCQDLP 533
Qy 375 -----LPVEKECEDHTNATLSDRGESLSPVSTRPV-----CSN 408
Db 534 PSQVSLPA-KALTEDTMSGGDLAATGDPAPAPPAFVPEVRLDSTYSQKAGAEQCS- 591
Qy 409 KLIDYILGGAPSDLEASSDSEEDWGEPEDDGFDSD--GSLSES-----DVEQDSEG- 459
Db 592 -----GDEEABAEABEVEEGEEDDEDESDNYGERSEAKSSMIETCGGEGG 643
Qy 460 --LHLWNSF---HSVDPPYKPNFTATIQTAARIAPRD---SDSGTSWSGSCGVGSCQE 510
Db 644 LSLRVQNSLRRRTHS-----EGSLLQEPGPGCFASDPTLHCSDGEGAASTWG 690
Qy 511 GPLPETPDHSSGE-----EDDWPSADEAENLKLWNSFCH 545
Db 691 MPSPFTLLKELGRNGSGMHLSLFTGHRKMSGADTVGDDDEASRKKKSNL----- 742
Qy 546 SEDPYNLL-----NFKAPFQSGKWKGRQDSKASSEVTAVFSGHHTLLSCKAQLLESQ 599
Db 743 AKQWKNKLGIFRRRNESPGAPPAGKADKMKSPKPTSSEALKW-----GESLEKL 792
Qy 600 EDNCPGCGLGEALAGERYTHIKRKVKYTFLEVEYIISGDEDRK 643
Db 793 LVHKYGLAVFQAFL-----RTEFSEENLEFWLACEDFKK 826

RESULT 10
PCT-US04-37982-202
; Sequence 202, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
; PRIOR FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 913
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
PCT-US04-37982-202

Query Match      3.7%; Score 138; DB 1; Length 913;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRARKPGPR-----LGSWFR-----LPFLRRSHACSESPPPP 38
Db 128 THGVARPEQRHSHLVCDSSDGLLGGWERYTEVAKRGQHTLPALSRATAPDNYII 187
Qy 39 SSRQNPNSALPERRTRYWTKLLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYS 98
Db 188 LAPLPGSQ-----LRLPVYQE-----DTIP----- 208
Qy 99 ALRASRGRESDAPTQKLSYTAAGLFAKTRVVSTLALARG----- 140
Db 209 -----BESGSPS--KGKSYT--GLGKSRMLKTVQTMKGHNQNCVVRPHATHSS 256
Qy 141 -GTPVAVLVRLEVKLKAQERALDSAPTFLLEQOLMGVELLPSSLQAGLVSHREL----- 195
Db 257 YGTVTTLAPKLVFPVFPVQPLDLCNPARTLLSBEILLYEGRNKAABVTLPAYSDLLFT 316
Qy 196 DSSSGPLSVQSLGNFKVSYLLNPVSYDYLPGLRQCSAGGQFVGRTLTTPESCYL 255
Db 317 KEDBPGRCDV-----LRNPLYQSV-----KLQEGSSDLKFCVL-----YL 353
Qy 256 SEDG-----CHPQPLR-----AEMSATAMRRCPPLSTEGLPETIHHRRMRWL----- 296
Db 354 AEKAECLFTLEAHSQEKRVKWCLENIAKQQQLADPDKMPETEADKREMALEBEGK 413
Qy 297 -----VFLQNGQDL-----PTLDQNGYHSLSEBHNLRMD----- 329
Db 414 PGAEDSPSKBPSGQBLPPGQDLPPNKPSPGQEPAPSPQEPPLSSKDSATSESGPPGPD 473
Qy 330 -----POHCTDNPAQAVSPAADRP---EPTKKPELVQE---VSQSPGSSLFCE--- 374
Db 474 PPSKDVPPCOEPPPAQDLSPCQDLPAQOEPLPHQDPLLTLDLPAIQESPTRDLPCCQDLP 533
Qy 375 -----LPVKECEBDHTNATDLSDRGESLPVSTRPV-----CSN 408
Db 534 PSQVSLPA-KALTEDTWSGDLAATGDPAPAPAFVPEVRLDSTYSQKAGAEQCS- 591
Qy 409 KLIDYILGAPSDLEASDSEDEWGEDDGFDS-GLSLES-----DVEQDSE- 459
Db 592 -----GDEDAEAEVEEGEGEDEDETDSDNYGERSEAKRSSMIETGQGAEG 643
Qy 460 --LHLWNSF-----HSVDPKPQNTATIQTAARAPRDP---SDSGTSMWSCGVSQCE 510
Db 644 LSLRVQNSLRRRTHS-----EGSLLOEPRGCFASDTTLHCSDGEGAATWG 690
Qy 511 GPLPETPDHSSGE-----EDDWEPSSADEAENLKLWNSFCH 545
Db 691 MPSPSTLKKELGRNGGSMHLSLFFTGHRKWSGADTVGDDDEASRKRKSNL----- 742
Qy 546 SEDPYNLL-----NFKAPFPQSGKWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQ 599
Db 743 AKDMKNKLGIFRRRNESPGAPPAGKADKMKWSFKPTSEALKW-----GESLEKL 792
Qy 600 EDNCPGCGLGEALAGERYTHIKRKYTFLEEVTEYVYISGDEDRK 643
Db 793 LVHKYGLAVFQAF-----RTEFSEENLEFWLACEDFKK 826

RESULT 11
PCT-US04-37982-192
; Sequence 192, Application PC/TUS0437982
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: CHIR0062-500 (23356.0003)
; CURRENT APPLICATION NUMBER: PCT/US04/37982
```

```
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 10/692,382
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: US 10/691,209
; PRIOR FILING DATE: 2003-10-21
; NUMBER OF SEQ ID NOS: 3504
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-37982-192

Query Match      3.7%; Score 138; DB 1; Length 1084;
Best Local Similarity 19.9%; Pred. No. 3.6;
Matches 164; Conservative 84; Mismatches 266; Indels 310; Gaps 34;

Qy 5 THRARKPGPR-----LGSWFR-----LPFLRRSHACSESPPPP 38
Db 299 THGVARPEQRHSHLVCDSSDGLLGGWERYTEVAKRGQHTLPALSRATAPDNYII 358
Qy 39 SSRQNPNSALPERRTRYWTKLLSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYS 98
Db 359 LAPLPGSQ-----LRLPVYQE-----DTIP----- 379
Qy 99 ALRASRGRESDAPTQKLSYTAAGLFAKTRVVSTLALARG----- 140
Db 380 -----BESGSPS--KGKSYT--GLGKSRMLKTVQTMKGHNQNCVVRPHATHSS 427
Qy 141 -GTPVAVLVRLEVKLKAQERALDSAPTFLLEQOLMGVELLPSSLQAGLVSHREL----- 195
Db 428 YGTVTTLAPKLVFPVFPVQPLDLCNPARTLLSBEILLYEGRNKAABVTLPAYSDLLFT 487
Qy 196 DSSSGPLSVQSLGNFKVSYLLNPVSYDYLPGLRQCSAGGQFVGRTLTTPESCYL 255
Db 488 KEDBPGRCDV-----LRNPLYQSV-----KLQEGSSDLKFCVL-----YL 524
Qy 256 SEDG-----CHPQPLR-----AEMSATAMRRCPPLSTEGLPETIHHRRMRWL----- 296
Db 525 AEKAECLFTLEAHSQEKRVKWCLENIAKQQQLADPDKMPETEADKREMALEBEGK 584
Qy 297 -----VFLQNGQDL-----PTLDQNGYHSLSEBHNLRMD----- 329
Db 585 PGAEDSPSKBPSGQBLPPGQDLPPNKPSPGQEPAPSPQEPPLSSKDSATSESGPPGPD 644
Qy 330 -----POHCTDNPAQAVSPAADRP---EPTKKPELVQE---VSQSPGSSLFCE--- 374
Db 645 PPSKDVPPCOEPPPAQDLSPCQDLPAQOEPLPHQDPLLTLDLPAIQESPTRDLPCCQDLP 704
Qy 375 -----LPVKECEBDHTNATDLSDRGESLPVSTRPV-----CSN 408
Db 705 PSQVSLPA-KALTEDTWSGDLAATGDPAPAPAFVPEVRLDSTYSQKAGAEQCS- 762
Qy 409 KLIDYILGAPSDLEASDSEDEWGEDDGFDS-GLSLES-----DVEQDSE- 459
Db 763 -----GDEDAEAEVEEGEGEDEDETDSDNYGERSEAKRSSMIETGQGAEG 814
Qy 460 --LHLWNSF-----HSVDPKPQNTATIQTAARAPRDP---SDSGTSMWSCGVSQCE 510
Db 815 LSLRVQNSLRRRTHS-----EGSLLOEPRGCFASDTTLHCSDGEGAATWG 861
Qy 511 GPLPETPDHSSGE-----EDDWEPSSADEAENLKLWNSFCH 545
Db 862 MPSPSTLKKELGRNGGSMHLSLFFTGHRKWSGADTVGDDDEASRKRKSNL----- 913
Qy 546 SEDPYNLL-----NFKAPFPQSGKWKGRQDSKASSEVTVAFGHHTLLSCKAQLLESQ 599
Db 914 AKDMKNKLGIFRRRNESPGAPPAGKADKMKWSFKPTSEALKW-----GESLEKL 963
Qy 600 EDNCPGCGLGEALAGERYTHIKRKYTFLEEVTEYVYISGDEDRK 643
Db 964 LVHKYGLAVFQAF-----RTEFSEENLEFWLACEDFKK 997
```


; CURRENT FILING DATE: 2005-04-29
; NUMBER OF SEQ ID NOS: 10858
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 94
; LENGTH: 3926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-675-841-94

Query Match 3.7%; Score 136.5; DB 8; Length 3926;
Best Local Similarity 23.9%; Pred. No. 24;
Matches 140; Conservative 68; Mismatches 198; Indels 179; Gaps 36;
Qy 61 LSQLALLPSLFQKLLWSQSGGLIPTRWLD--FAASY-----SALRA 102
Db 2779 LSQLVSRQPPKSPQ-VLYSPVS-PLSPHRLDTSFASSELNKAHVSPQKHFTADSAIR- 2835
Qy 103 SRGRESADP--TVQKSL-----YTAAGLFAKTRVVSTLALARGGTPVAVLVRLEVKL 155
Db 2836 ---QOTLPRPMKTLQRLSDPKPLSPTABESAKERF--SLYQHGGGLGSQVSALPPNSLV 2890
Qy 156 KAQERALDSAAP-----TFLLQQQLMGVVELLPSSLOAGLVSHRELDSS 198
Db 2891 RKVKRTLPSPPEEAHLPLAGQASPOLYAAASLLQRLGTGTTVPAT--KASLL--RELDL- 2946
Qy 199 SSGPLSVQSLGNFKVSYLLN-----PSYDLYLPQLGL--RCQSSA--- 237
Db 2947 -----DLRLVEHSTKLKQKQAEIDEEKEIDAKLYL-ELGITQKESLAKDR 2994
Qy 238 GGGQFVGFRITLPESCYLE-----DGC-HPQPLRAEMSATAWRCPPPLSTEGULPEI 288
Db 2995 GGRDYPPLRGLGHRDYLSDSELNQLRLQCGCTTPAGQVDFPATA---AAPATPSGPTAF 3051
Qy 289 HHRMRWLVPLOP-----NOQDLPTLDQDNGVHSLSEHNLLRMDPOHCTDN----- 336
Db 3052 QQPR-----FQPPAPQYSAGSGGPT--QNGF-----PAHQAPTYPGPSTY 3089
Qy 337 PAQAVSPAADRPEPTEKKPELVIQE-----VSQSPQ---GSSLFCELPVEKECEDH 385
Db 3090 PAPAPPGASYP---AEPGLPNQQAFRPTGHYAGQTPMTTQSTLF-PVPADSRAPLQK 3144
Qy 386 TNATDLSDRGESLPVSTRPVCNKLIDYILGGAPSDLEASDSESDW--GEEPEDDGFD 443
Db 3145 PRQTSADLEQKVPTNYEVIASPVV---PMSSAPSETSYSGPAVSSGYEQGVPEVPRAG 3201
Qy 444 SDGSLSDVEQDSEGLHLMNSFHSVDPYKPNF-----TATIQTAARIAPR 490
Db 3202 DRGSVQSAPTYPSDSH-----YTSLEQNVPRNYVMIDDISETLTKDSTSTAPDSQRLEPL 3257
Qy 491 DPDSGTSWS--GSCGVGSCQEGP-LPETPDHSSGEEEDDWEPSAD 532
Db 3258 GPGSSGRPGKEPGEVPGV---LDGPTLPCC--YARGESEEDSDYD 3297

RESULT 15

US-60-710-726-316
; Sequence 316, Application US/60710726
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Serre, Valerie
; APPLICANT: Eardewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: GeneMap of the human genes associated with Crohn's disease
; FILE REFERENCE: 059908-5009-PR

; CURRENT APPLICATION NUMBER: US/60/710,726
; CURRENT FILING DATE: 2005-08-24
; NUMBER OF SEQ ID NOS: 20166
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 316
; LENGTH: 3926
; TYPE: PRT
; ORGANISM: Homosapiens
US-60-710-726-316

Query Match 3.7%; Score 136.5; DB 8; Length 3926;
Best Local Similarity 23.9%; Pred. No. 24;
Matches 140; Conservative 68; Mismatches 198; Indels 179; Gaps 36;
Qy 61 LSQLALLPSLFQKLLWSQSGGLIPTRWLD--FAASY-----SALRA 102
Db 2779 LSQLVSRQPPKSPQ-VLYSPVS-PLSPHRLDTSFASSELNKAHVSPQKHFTADSAIR- 2835
Qy 103 SRGRESADP--TVQKSL-----YTAAGLFAKTRVVSTLALARGGTPVAVLVRLEVKL 155
Db 2836 ---QOTLPRPMKTLQRLSDPKPLSPTABESAKERF--SLYQHGGGLGSQVSALPPNSLV 2890
Qy 156 KAQERALDSAAP-----TFLLQQQLMGVVELLPSSLOAGLVSHRELDSS 198
Db 2891 RKVKRTLPSPPEEAHLPLAGQASPOLYAAASLLQRLGTGTTVPAT--KASLL--RELDL- 2946
Qy 199 SSGPLSVQSLGNFKVSYLLN-----PSYDLYLPQLGL--RCQSSA--- 237
Db 2947 -----DLRLVEHSTKLKQKQAEIDEEKEIDAKLYL-ELGITQKESLAKDR 2994
Qy 238 GGGQFVGFRITLPESCYLE-----DGC-HPQPLRAEMSATAWRCPPPLSTEGULPEI 288
Db 2995 GGRDYPPLRGLGHRDYLSDSELNQLRLQCGCTTPAGQVDFPATA---AAPATPSGPTAF 3051
Qy 289 HHRMRWLVPLOP-----NOQDLPTLDQDNGVHSLSEHNLLRMDPOHCTDN----- 336
Db 3052 QQPR-----FQPPAPQYSAGSGGPT--QNGF-----PAHQAPTYPGPSTY 3089
Qy 337 PAQAVSPAADRPEPTEKKPELVIQE-----VSQSPQ---GSSLFCELPVEKECEDH 385
Db 3090 PAPAPPGASYP---AEPGLPNQQAFRPTGHYAGQTPMTTQSTLF-PVPADSRAPLQK 3144
Qy 386 TNATDLSDRGESLPVSTRPVCNKLIDYILGGAPSDLEASDSESDW--GEEPEDDGFD 443
Db 3145 PRQTSADLEQKVPTNYEVIASPVV---PMSSAPSETSYSGPAVSSGYEQGVPEVPRAG 3201
Qy 444 SDGSLSDVEQDSEGLHLMNSFHSVDPYKPNF-----TATIQTAARIAPR 490
Db 3202 DRGSVQSAPTYPSDSH-----YTSLEQNVPRNYVMIDDISETLTKDSTSTAPDSQRLEPL 3257
Qy 491 DPDSGTSWS--GSCGVGSCQEGP-LPETPDHSSGEEEDDWEPSAD 532
Db 3258 GPGSSGRPGKEPGEVPGV---LDGPTLPCC--YARGESEEDSDYD 3297

Search completed: September 16, 2005, 10:56:23
Job time : 86.5911 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 08:56:27 ; Search time 17.8086 Seconds
(without alignments)
3771.163 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRKRKPGRLGWSF.....MFNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR_79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	5.5	657	2 S10001	MyD116 protein - m
2	171.5	4.6	590	2 A56535	gadd34 protein - l
3	151.5	4.1	1510	2 T33100	hypothetical prote
4	141	3.8	522	2 T40520	hypothetical prote
5	136.5	3.7	1571	2 T00062	hypothetical prote
6	135	3.6	527	2 C88042	protein P56D12.4 [
7	133.5	3.6	782	2 T32155	hypothetical prote
8	133.5	3.6	883	2 S49136	brevican precursor
9	132.5	3.6	682	2 P90603	vipe-like (mycopla
10	130	3.5	613	2 T42671	hypothetical prote
11	130	3.5	1560	2 T30282	calcium-binding pr
12	129.5	3.5	1473	2 T31422	C-terminal domain-
13	127.5	3.4	406	2 S38170	SRP40 protein - ye
14	126	3.4	817	2 T03852	protein phosphatas
15	126	3.4	1229	2 T46116	hypothetical prote
16	123.5	3.3	4377	2 A55575	ankyrin 3, long sp
17	123	3.3	1324	2 T01508	mismatch repair en
18	123	3.3	4549	2 T20771	hypothetical prote
19	123	3.3	4667	2 T20774	hypothetical prote
20	122	3.3	1189	2 S56852	hypothetical prote
21	121.5	3.3	953	2 T01093	luminidependens pr
22	120.5	3.2	306	1 BVBYK6	MAK16 protein - ye
23	120.5	3.2	660	2 S31437	homeotic protein H
24	120	3.2	1684	2 JW0057	gravin - human
25	119.5	3.2	961	2 A55380	faciogenital dyspl
26	119.5	3.2	1316	2 T00381	KIAA0633 protein -
27	119.5	3.2	1325	2 S62497	probable nucleopor
28	119.5	3.2	1468	2 S11515	formin - mouse
29	119.5	3.2	3942	2 T42730	Bassoon protein -

30	118.5	3.2	313	2 G84413	hypothetical prote
31	118	3.2	1026	2 T28968	hypothetical prote
32	118	3.2	1239	2 G02750	DNA-directed DNA p
33	118	3.2	1829	2 T26135	hypothetical prote
34	118	3.2	4342	2 H83343	probable non-ribos
35	117.5	3.2	893	2 S57653	brevican precursor
36	117.5	3.2	925	2 T48834	hypothetical prote
37	117.5	3.2	5762	2 A41819	proline-rich pepti
38	117	3.2	184	2 A42749	hypothetical prote
39	117	3.2	1231	2 S30185	insulin receptor s
40	117	3.2	1317	2 T03748	hypothetical prote
41	117	3.2	1341	2 T17285	apoptosis associat
42	116.5	3.1	912	2 A54423	brevican precursor
43	116.5	3.1	1331	2 AE1843	hypothetical prote
44	116	3.1	699	2 I38073	nucleolar phosphop
45	116	3.1	1300	2 T03166	probable immediate
46	116	3.1	1974	2 T16703	hypothetical prote
47	115.5	3.1	1621	2 A82255	hypothetical prote
48	115	3.1	1241	2 JU0466	potassium transpor
49	114.5	3.1	860	2 T39502	hypothetical prote
50	114.5	3.1	1163	2 A36685	205K microtubule-a
51	114.5	3.1	1210	2 I39410	AF-4 protein, spli
52	114.5	3.1	1213	2 A58198	serine/proline-ric
53	114.5	3.1	1485	1 ISZPT2	DNA topoisomerase
54	114.5	3.1	2524	2 A35844	Xotch protein - Af
55	114	3.1	1463	2 T30290	AAS surface protei
56	114	3.1	2437	2 S53611	MBP1 protein - ra
57	113.5	3.1	699	2 AC0504	Probable bifunctio
58	113.5	3.1	897	2 E65127	hypothetical prote
59	113.5	3.1	1180	2 T20773	transforming prote
60	113	3.0	419	1 TVLWMC	hypothetical prote
61	113	3.0	823	2 T21104	hypothetical prote
62	113	3.0	1495	2 T48429	hypothetical prote
63	113	3.0	2009	2 S49764	SEC7 protein - yea
64	112.5	3.0	382	2 S72169	DNA-binding protei
65	112.5	3.0	681	2 P85062	hypothetical prote
66	112.5	3.0	1805	2 A34736	nestin - rat
67	112	3.0	1185	2 A42404	collagen adhesin -
68	112	3.0	1237	2 A34598	ecdysone-induced p
69	112	3.0	1394	2 B34598	ecdysone-induced p
70	112	3.0	3938	2 T42761	Bassoon protein -
71	111.5	3.0	364	1 EDBESP	immediate-early pr
72	111.5	3.0	619	2 T08613	hypothetical prote
73	111.5	3.0	757	2 S68142	probable transcrip
74	111.5	3.0	1249	2 T14150	vesicle associated
75	111.5	3.0	3707	2 S18252	heparan sulfate pr
76	111	3.0	549	2 T40692	hypothetical prote
77	111	3.0	1235	1 S16948	insulin receptor s
78	111	3.0	1791	2 T02345	hypothetical prote
79	110.5	3.0	764	2 JC5113	ribosomal transcri
80	110.5	3.0	764	2 JC5112	transforming prote
81	110.5	3.0	906	2 A43817	A-alpha 2 4 protei
82	110.5	3.0	930	2 D37271	hypothetical prote
83	110.5	3.0	932	2 T48761	mdm2 protein - mou
84	110	3.0	489	2 S13349	probable membrane
85	110	3.0	509	2 T34871	hypothetical prote
86	110	3.0	649	2 T46500	hypothetical prote
87	110	3.0	816	2 T17257	hypothetical prote
88	110	3.0	1487	2 T02850	alpha-A-crystallin
89	110	3.0	2688	2 T49477	hypothetical prote
90	109.5	3.0	842	2 T23715	hypothetical prote
91	109.5	3.0	1062	2 T46444	hypothetical prote
92	109.5	3.0	1093	2 I38533	AF17 protein - hum
93	109.5	3.0	1231	2 T18532	serine/threonine p
94	109.5	3.0	1242	2 JS0670	neulin receptor s
95	109.5	3.0	1286	2 T23714	hypothetical prote
96	109.5	3.0	1596	2 A35927	190K DNA-binding p
97	109.5	3.0	1624	2 T25592	hypothetical prote
98	109.5	3.0	1757	2 T05204	hypothetical prote
99	109.5	3.0	2774	2 A43359	microtubule-associ
100	109.5	3.0	3924	2 S37431	ankyrin 2, neurona

ALIGNMENTS

RESULT 1
S10001

MyD116 protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C:Accession: S10001
R:Lord, K.A.; Hoffman-Liebermann, B.; Liebermann, D.A.
Nucleic Acids Res. 18, 2823, 1990
A:Title: Sequence of MyD116 cDNA: a novel myeloid differentiation primary response gene
A:Reference number: S10001; MUID:90251472; PMID:2339071
A:Accession: S10001
A:Molecule type: mRNA
A:Residues: 1-657 <OR>
A:Cross-references: UNIPROT:P17564; EMBL:X51829; NID:G53040; PIDN:CAA36128.1; PID:G53041

```
Query Match          5.5%; Score 204; DB 2; Length 657;
Best Local Similarity 21.5%; Pred. No. 4.9e-06;
Matches 149; Conservative 86; Mismatches 290; Indels 168; Gaps 27;

Qy 60 LLSQALLPSLFQKLLWSQLGLIPTRWLDFAASYALRASGRBESDAPTVQKSL 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 LLSPLMGLLSR-----AWSRLRGPEVPEAWL-----AKTVTGADQ 55

Qy 120 YTAAGLFAKTRVVSTLALARG-----GTPVAVLVLRLEVKLKAQERAL---DSAAPTFL 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 IEAALPTPTVSGNL-LPHGETESGSP-----EQQAQRKCLVEAESPPP----- 102

Qy 172 EQQLMGVELLPS-SIQAGLVSHRELD-SSSGPLSVQSLG-----NFKVVSVY---LLN 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 --ETWGLSNVDVYNAKPQDDLRKEMERTAGKATLQAGQADKRLGEVVAEEGVAE 160

Qy 220 PSYLDYLPQLGLRCOSSAGGQGVGFRYLTPTESCVELSDGCHPQPLRAEMSATWRRCPP 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 PAY-----PTSQLEGPAEN-----EDGETVKTYQASAASIAFGYKPS 199

Qy 280 LSTEGLPEIHRRMRWLVPQNQODLPTLDQNGVHSLSEHNLLRMDPHQCTDNP 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 TPVPFLGAEHQATE-----EKGTEKADPSNPSGGSHRAWAY-YREKPKQSGEAKVE 254

Qy 340 AVSPAADRP-----EPTKKPLVIOEVQSQSGSLFCPLPVEKECEDHTNATDLSDRG 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 AHRAGQGHPCRNAEAEEGPETTF--VCTGNAFLKAWYRPGEDTEEDNDSDSABEDT 312

Qy 396 ESLPVSTRPVCSNKLIDYILGAPSDLEASDSSESD-----W-----GE 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 AQTGATPHTSAFLKAWYRPGEDTEEDSDSABEDTAQTGATPHTSAFLKAWYRPG 372

Qy 436 EPEDDGFDGSLSDVEQDSEGLHLWNSFHSVDPYPKQNFATIQTAARIAPRDP 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 DTEENSOLDL--AEEDTAQTGATPHT--SAFLKAWYRPGEDTEEDNDSABEDTAQT 429

Qy 496 GTSWSGSCGVSCQGGPLPETPDHSSGEEDDWEPSDAENLKLWNSFCHSDPYNL 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GATPHTSPFLKAWYRPGEDTEDTEEDSENVAPOGSETADSSQSPC----- 478

Qy 556 KAPFQPS-----GKNWKGQDSKASSEVTVAFGHH-----TLSCKAQLLS-- 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 ---LQPRCLPGEKTKGRGEPPLFQVAFYLFGEKPSFWAAPKLPLRLQRLRLFKAPT 535

Qy 599 --QEDNCPGCGLGEALAGERYTHIKKKVTLFEEVTEYIS-----GDEDKGPWEERFAR 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 RDQPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPEWFOFAR 580

Qy 652 DGCGRFQKRIQTEVAIGYCLAPEHREKMFNRLR 684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 DRSRFARIAQAEKLGPLYLPDSEARAWARLR 613
```

RESULT 2
A56535

gadd34 protein - long-tailed hamster
N:Alternate names: MyD116 protein homolog
C:Species: Cricetus longicaudatus (long-tailed hamster)
C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C:Accession: A56535
R:Zhan, Q.; Lord, K.A.; Alamo Jr., I.; Hollander, M.C.; Carrier, F.; Ron, D.; Kohn, K.W.
Mol. Cell. Biol. 14, 2361-2371, 1994
A:Title: The gadd and MyD genes define a novel set of mammalian genes encoding acidic pr
A:Reference number: A56535; MUID:94187707; PMID:8139541
A:Accession: A56535
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-590 <ZHA>
A:Cross-references: UNIPROT:Q60465; GB:L28147; NID:G452489; PIDN:AAA36983.1; PID:G452490
C:Genetics:
A:Gene: Gadd34

```
Query Match          4.6%; Score 171.5; DB 2; Length 590;
Best Local Similarity 18.1%; Pred. No. 0.00069;
Matches 100; Conservative 68; Mismatches 198; Indels 187; Gaps 18;

Qy 250 PESCYLSDGCHPQPLRAEMSATWRRCPP-----STEGLPETIHHRR 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 PETLGLSDDDKQGGQGPREGQGRAHTAGLPIILLSFGLQSADKSLGEVVAGEGVTETAYPT 162

Qy 293 MRWLVLFPQNGODLPTLDQNG-----YHSLSEEH 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 SHW-----EGCPSEEDGETVKAFRASADSPGHKSSSTSVYCPGAHQATEKQ 213

Qy 324 NLLRMDP-----QHCTDNPAAQVSPAADREP-----TEKKPELV 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 TENKADPPSPSGSHRAWYCSKQGEA-----DPEHRAGKYQLCQNAEAESEAK 267

Qy 359 IOEVQSPOQSSL--FCPLPVEKECEDHTNATDLSRGESLPVSTPVCNKLIDYILG 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 VSSLVSSGNAFLKAWYRPGEDTEDDDSDGSAEEGKALSSPTSPE-----HDFLKA 322

Qy 417 GAPSDLEASDSSESDGERPEDDGFSDGSLSDVEQDSEGLHLWNSFHSVDPYPKQN 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 WYTRPGEDTEDDDSDGSAEE-----EGKALSSPTSPEHDFLKA-----VYRPG 369

Qy 477 FTATIQTAARIAPRDPDSGTSGSGCGVSCQGGPLPET----- 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 DT-----EDDQSDWGSAEKQGLAQTFATPHTSAFLKTVWCPCGEDTEDDDC 416

Qy 517 ----PDHSSGEEDDWEPS-----ADEAENLKLWNSFCHSDPYNLLNFKAPFPQSGK 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 EVVVPEDSEAADPKSPSHEAQGLPGEQTEGL-----VEAHSFLFOVAYFLPGEKAP 470

Qy 565 NWKGRQDSKASSEVTVAFGHHHTLLSCKAQLLSQEDNCPGCGLGEALAGERYTHIKKK 624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 PW-----TAPKLPLRLQRLTLRLTPTQ---DQDPETP-----LRARK 505

Qy 625 VTFLFEEVTEYIS-----GDEDKGPWEERFARDCGRFQKRIQTEVAIGYCLAPEHREKM 679
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 VHFSENVTVHFLAVWAGPAQAARRGPEWQLARDRSRPARRIAQAEKLGPLYLTPAFRARA 565

Qy 680 FNRIRIESKDLLL 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 WARLGNPSLPLAL 578
```

RESULT 3

T33100

hypothetical protein H1LE01.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T33100

R:Geisel, C.; Wamsley, P.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid H1LE01.

A:Reference number: Z21282

A:Accession: T33100

R;Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, M.; Ohira, M.; Miyajima, N.; Tanaka, A.; submitted to the EMBL Data Library, October 1997
A;Description: Prediction of the coding sequences of unidentified human genes. VIII. The A;Reference number: Z14082
A;Accession: T00062
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-1571 <ISH>
A;Cross-references: UNIPROT:O43161; EMBL:AB007894; NID:G2662148; PIDN:BAA23707.1; PID:G262148
A;Experimental source: brain; clone HH2165
C;Genetics:
A;Note: KIAA0434

```
Query Match          3.7%; Score 136.5; DB 2; Length 1571;
Best Local Similarity 23.9%; Pred. No. 0.66;
Matches 140; Conservative 68; Mismatches 198; Indels 179; Gaps 36;

Qy 61 LSQALLPSLFQKLLWSQSGGLIPTRWLD--FAASY-----SALRA 102
Db 424 LSQVSRQPPKSPQ-VLYSPVS-PLSPHRLDTSFASERLNKAHVSPQKHTADSALR- 480

Qy 103 SRGRESAP--TVQXSLS-----YTAAGLFAKTRVSTLALARGTTPVAVLRLVKL 155
Db 481 ---QOTLPRPKMTLQSLSDPKLSPTAESAKERF--SLYQHOGGLGSQVSALPPNSLV 535

Qy 156 KAQRALDSAP-----TFLEQQLWGVLLPSSLOAGLVSHRELDSS 198
Db 536 RKVKRLTSPPEEHLPLAQASPOLYAAQLLQGLGTPTVPAT-KASLL--RELD- 591

Qy 199 SSGPLSVQSLGNFKVSYLLN-----PSYLDYLPQLGL--RCOSSA--- 237
Db 592 -----DLRLVEHESKLRKKQAEDEBEKEIDAKLYL-ELGITQKESLAKOR 639

Qy 238 GGGQVGVGRTTTPSCYUSE-----DGC-HQOPLRAEMSAWRCRPPPLSTEGUPEI 288
Db 640 GGRDYPPLRGLGHRDYLSDSELNQLRQGGCTTPAGQVDFPATA--AAPATPSGPTAF 696

Qy 289 HHRMRWLVLQP-----NQODLPTLDQDQYHSLSEEHLLRMDPOHCTDN----- 336
Db 697 QQPR-----FQPPAPQVSAGSGPT-----PAHQAPTPGPGSTY 734

Qy 337 PAQAVSPAADRPETKEKPELVIQ-----VSQSPQ-----GSSLFCELFVEKECEDH 385
Db 735 PAPAPPPGASYP-----AEPGLPNQAQFRTPGHYAGQTPMTTQSTLFPVPAADSRAPLOK 789

Qy 386 TNATDLSDRGSLPVSSTRPVCSNKLIDYLGAPSDLEASDSSESDW--GERPEDDGF 443
Db 790 PROTSADLEQKVPNTYEVIASPVV---PMSSAPSETSYSGPFAVSSGYEQGVPEVPRAG 846

Qy 444 SDGSLSESDVQDSBGLHLMNSFHSVDPYKQNF-----TATIQTAARIAPR 490
Db 847 DRGSVSGPAPTYPSDSH-----YTSLEQNVPRNYVMIDDISELTKDSTSTAPDSQRLEPL 902

Qy 491 DPSDSGTSWS--GSCGVGSCQEGP-LPPTPDHSSGEEDDWEPFAD 532
Db 903 GPGSSGRPGKEPGBGV---LDGPTLPC--YARGESESESDY 942
```

RESULT 6
C88042
protein F56D12.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: C88042
R;anonymus, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A;Reference number: A75000; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/errata
A;Accession: C88042
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-527 <STO>
A;Cross-references: GB:chr_II; PIDN:AB66118.1; PID:G2315675; GSPDB:GN00020; CESP:F56D12.4
C;Genetics:
A;Gene: F56D12.4
A;Map position: 2

```
Query Match          3.6%; Score 135; DB 2; Length 527;
Best Local Similarity 21.3%; Pred. No. 0.18;
Matches 112; Conservative 57; Mismatches 178; Indels 180; Gaps 27;

Qy 96 SYSA-----LRASRGRESDAPTQKSLSYTAAGLFAKTRVSTLALARGTTPVAVLRL 150
Db 33 SYSADWLHNDNSGGDDSEEDSABFLVHPCLRSIIKSK--SAYAVPRNSDDVAAAAAM 90

Qy 151 LEVLKQAE-----RALDSAAPTFLRQQLWGVLLPSSLOAGLVSHRELDSSSGP 202
Db 91 KNNRIKPSKTSFEISLKMMESTSKSABEAESVADLLPT-----LDEKWPDP 139

Qy 203 LSVQSLGNFK-----VVSYLNPVSYLDYLPQLGLRCQSSAGGQFVGFRTLTTPESCYL 255
Db 140 --EENENLKSRRKSNIMRY--SPSLCDSDLSLNLHDKSDEDEDRLNR--PSGCWS 192

Qy 256 SEDGCHPOPLR--AEMSAWRCRPPPLSTEGLPPIHHRMRWLVLQPNQODLPTL--- 310
Db 193 SPNFFDETFRGVSPMNQKCM-----SAPLEDIVFKTCNVLYNDYDVIGGKIPILREL 246

Qy 311 -DQDN---GYHSLSEEHLLRMDPOHC-----TDNPAQAVSPAADRPETKEK 354
Db 247 IDSDIAIGYFVETPDN---DPSVCSYRLPSGLSAPSPRPRLSPVFGQAPPQYR 302

Qy 355 PELVIQVSOS---PQSSSLFCELFVEKECEEDHTNATDLSDRGESLPVSTRPVCSNKL 411
Db 303 -----MVVSQSMHLPRPDIEAF---ADQNDYQ----- 330

Qy 412 DYILGAPSDLEASDSSESDWGEEDDGFSDGSLSESDVEODSGLHLMNSFHSVDP 471
Db 331 -----APPTFNIGSDLEDNITEEBEDBED-----EEDVRKTEE---VW----- 367

Qy 472 YKQNFAT-----IQTAA-----RIAPRDPDSGTSWSGSCGVGSCQEGPLP--- 514
Db 368 --PENATANFMSDGVRRVHRSAFTQINRYLDDDDVTDGGT-----VSPPLPRIT 414

Qy 515 -----ETPDHSSGEEDDWEPFSADEAENLKLWNSFCHS 546
Db 415 VSRSAHSQLYNLGAETEDDADGGDD---DDDVAENL---NENCSS 455
```

RESULT 7
T32155
hypothetical protein C10F3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T32155
R;Davidson, S.; Wohldmann, P.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid C10F3.
A;Reference number: Z21127
A;Accession: T32155
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-782 <DAV>
A;Cross-references: UNIPROT:O16883; EMBL:AF022968; PIDN:AAB69883.1; GSPDB:GN00023; CESP:
A;Experimental source: strain Bristol N2; clone C10F3
C;Genetics:
A;Gene: CESP:C10F3.1
A;Map position: 5
A;Introns: 52/3; 107/2; 224/1; 318/2; 378/2; 455/1; 657/2

```
Query Match          3.6%; Score 133.5; DB 2; Length 782;
Best Local Similarity 20.9%; Pred. No. 0.4;
Matches 108; Conservative 67; Mismatches 188; Indels 153; Gaps 23;

Qy 222 YLDYLPQLGLRCQSSAGGQ--FVGFRTLTPESCYLESDGCHGPQLRAEMSAWRCRCPPL 280
```

```
Db 285 WADMLANIGQPPSPQDAQGFENFROFTNDLCRIGD---CMLDCIRSKFNT-----RCEGS 338
Qy 281 STEGLPIHHR---RMRWLVLQPNQGDLPDLDQDNGYHSLSEHNLLRMD----- 329
Db 339 AGTLLSVFVRPIAATQNKLSILRPIIGTWP---EOCGYLTNNAELKKHRIIDATWDEEL 395
Qy 330 -----POHCTDNPAQ-----AVSPAADRPEPEKK-PELVIQE----- 361
Db 396 KRWYAEKIAKEARDTAQDEILANLVPLDENGVLPRALPELKSIESPLDVSVKTLDLQLI 455
Qy 362 -----VSQSPQSSSLFCELPVKECED-----HTNATDLSDRGESL 398
Db 456 LDMYSNNKTBELNISEKNVTSFSE-PSEKEDASTVTSVSLHTNATD----- 506
Qy 399 PVSTRPVCSNKLIDYI-----LGGAPSDLEASSD--SESEDKGEBEPDDG- 441
Db 507 -----SEILEHISEKSTEESSGSGSGEMSGDGSNEASGEGSGYDASGSGDNGSE 557
Qy 442 FDSGSLSESDVEODSEGLHLMNSFHSVDPYKPNQNTATTQTAARIAPRDPDSGTSWSG 501
Db 558 FNSGSGSEASEGEGSGSDQSGN---YK---MIESIESGEPFSGSGSGSGDTPASS 610
Qy 502 S-----CGVSCQEGPLPETPDHSSGER-DDWEPADAEANLKLNSFCHSEDPY 550
Db 611 DTSIDDKSIIRSGESAES--VSEILOEASGEDAPTTLPTSESTGYKIDHS----- 660
Qy 551 NLLNFKAPFQSGKNWKGRODSKASSVTVAFSGHHTLLSCAKQLLESQDNCNCPG----- 605
Db 661 ---GFGSSSGSGSIELRDSGEGSAEYDASGSGD-----NSGDFNSSGSGSGEA 707
Qy 606 CGLGEALAGBRYTHIRKKTVEEYVYISGDED 641
Db 708 SGVSGSGSDQSGNGYKIEVIESGSDYBFGSSN 743

RESULT 8
brevican precursor - rat
N;Alternate names: aggrecan-like protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 01-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C;Accession: S49126; 155457
R;Seidenbecher, C.I.; Langnase, K.; Wax, H.; Seidel, B.; Garner, C.C.; Gundelfinger, E.D.
submitted to the EMBL Data Library, June 1994
A;Description: Molecular cloning of a new member of the aggrecan/versican family of proteoglycans
A;Reference number: S49126
A;Accession: S49126
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-883 <RES>
A;Cross-references: UNIPROT:P55068; EMBL:X79881; NID:G509396; PIDN:CAA56255.1; PID:G509393
R;Seidenbecher, I.C.; Richter, K.; Rauch, U.; Fassel, R.; Garner, C.C.; Gundelfinger, E.
J. Biol. Chem. 270, 27206-27212, 1995
A;Title: Brevican, a Chondroitin Sulfate Proteoglycan of Rat Brain, Occurs as Secreted and Membrane-associated Forms
A;Reference number: 155457; MUID:96070828; PMID:7592978
A;Accession: 155457
A;Status: translated from GR/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-883 <RES>
A;Cross-references: EMBL:X79881; NID:G509396; PIDN:CAA56255.1; PID:G509397
C;Comment: For an alternative splice form, see PIR:A53908.
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF
C;Keywords: alternative splicing
F;49-138/Domain: immunoglobulin homology <IMM>
F;173-250/Domain: link protein repeat homology <LNK1>
F;271-352/Domain: link protein repeat homology <LNK2>
F;626-657/Domain: EGF homology <EGF>
F;664-784/Domain: C-type lectin homology <LCH>
F;791-847/Domain: complement factor H repeat homology <FHD>
```

Query Match 3.6%; Score 133.5; DB 2; Length 883;
Best Local Similarity 22.2%; Pred. No. 0.47;

```
Matches 154; Conservative 63; Mismatches 217; Indels 259; Gaps 40;
Qy 140 GGTPTAVLV---LRLEVKLKAQERALDSAAFTLLEQLQMGVELLP---SSLQAGLVSHRE 194
Db 83 GDREVEVLVARGLVKVNAYRFRVALPAYPASLTDVSLVLSLRPNDSGVYRCEVQHG 142
Qy 195 LSSSSGPIUSVQSL-----GNFKVSVYLLNFSYLDYLPQLGLRCQ 234
Db 143 DSSDAVEYKVGWVFLYREGSARYAFSPAGQACARIGARIATPEQL-YAAYLGGYEQ 201
Qy 235 SSAG--GGQFVGFRITLP--ESCVLSEDG-----CHPOPLRAEMSA- 271
Db 202 CDAGWLSQDQTVRIPIQNPRACGDMGDPGVGVNYGVGPDLDYDVCYAEADLNGELFLG 261
Qy 272 -----TAMR-----RCP-----PLSTE----- 283
Db 262 APPGKLTWBEARDYCLERGAQIASTQGLYAAWNGGLDRCSPGWLADGVSRYPIITPSQRC 321
Qy 284 --GLPIHHRMRWLVLQPNQGDLPDLDQDNGYHSLSEHNLLRMDPOHCTDNPA--- 338
Db 322 GGGLPVK-----TLFLPNQ-----TGPFSSKQNRFNV-----YCFRDSAHPS 359
Qy 339 ---QAVSPAADRPEP---TEKKPELVI-QEVSQSPQSGSLFCELPVKECEDEHTNATD 390
Db 360 AFSEASSPASDGLAIVTTEKLEELQLQEAVERSESRAIY-SIFI-----TE 407
Qy 391 LSDRGESLPVSTRPVCNKLIDYLGAP--SDLEASSDS-----ESEDWGEPEDDGDFD 443
Db 408 DGGGSGSTEDP-----AEAPRTPLESETQSVAPPTGSSERSEGEALEEERF 454
Qy 444 SDGSLSESDVEQSEGLHLW-----NSFHSVDPYKPNQNTATTQTAARIAPRD 491
Db 455 KDTETPKKEEQ--ENLWVMTLSPLPTGLEHSLSQVSP--AAVLQVQASPSRP 511
Qy 492 PSDSGTSSGSCGVSQCO---EGPLPETDHS---SGEEDDWEPSEDAEENLKLNSFCH 545
Db 512 PRVHGPT-----VETLQPPGSGSLSTPDGAREVGGTGPSELGVPFREREAGSS--S 563
Qy 546 SEDPYNLLNFKAPFQPSG-----KNMKGR-----QDSKASSEVTVA 582
Db 564 LEDGPSLL--PETWAPVGTREVETPSEKSGRTVLGTSVQAQPVLPDTSASRGGVAVAP 621
Qy 583 SGHTTLLS-CK--AQLESQED-NC---PG-----CGLG-----EALAGERYTHIKR 622
Db 622 SSGDCIPSPCHNGTCLBEKEGPRCLCVPGYGGDLCDVGLHFCSPGWEPFQGCYKHFST 681
Qy 623 KKVTFLEEVEYVYISGDEDRKGWPWFARDGCR 655
Db 682 RR-----SWEE-AESQCR 693
```

RESULT 9

P90603
vipe-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAB C)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: F90603
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: F90603
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-682 <KUR>
A;Cross-references: UNIPROT:Q98PI8; GB:AL445566; PID:gl4090149; PIDN:CAC13907.1; GSPDB:GSPDB:G
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU 7340
A;Genetic code: SGC3

Query Match 3.6%; Score 132.5; DB 2; Length 682;
Best Local Similarity 20.6%; Pred. No. 0.39;

Db 1137 --VSEG-----YVSESGQSESESESEGESEGESETEG-----AEE 1171

Qy 337 PAQAVSPAADRPETPE--KKPELVIOEVSVSPQSSSLFCPLPVKEKECEEDHTNATDLSDR 394

Db 1172 PHEQMESEGGQOESEGAEPQOEQMESEGGEP-----ESEDVDSGEITDGSMD 1220

Qy 395 GESLPVSTRPVCSNKLIDYILGGAPSDI-----EASSDSESDWGEPEDDG-----F 442

Db 1221 NMPQFGSSD-----GGVPLGLTEVDEQTSDGEEP---EPEEPGEGQGPHEQM 1265

Qy 443 DSDGSLSESDVEQ-----DSGLHLWNSFHSDVP-----YKPNQNFATI 481

Db 1266 ESEGEHSEHEVEEPQEVIESEG-----QQSESEDPFEPGEPESEGEQEPQEWESI 1321

Qy 482 ----QTAARI---APRDPDSGTSMWSGSC----- 503

Db 1322 VQEPESGADVDSGEVTDGSDNMMPQGSDDGAGGVPLGLVEVKTNSQEGSEVQESQDPQ 1381

Qy 504 -----GVGSCQEGPLPETPHSHSGEEDDWEPSDAEENLKLWNSFCHSDSDPPNLLNFKA 557

Db 1382 EQMEGEGQSEGESEPEEGEQGPQOEVESEGESEEE-----EVEEPQGVMESEG 1436

Qy 558 PFQPSGK-NMKGRQDSKASSVTVAFSGHHTLLSCKAQLLESQED-----NCP 604

Db 1437 QOSESEPEFEGEPSESEGE-----QEPQOMSEGEQEPSEEDVDGSEVTDGSDNMMP 1491

Qy 605 GCGLGEALAGERYTHIKRKKVTFLEEVTEYYISGDEDRKGPWEFFARDGCRFQRIQETE 664

Db 1492 GQGSDDGAGGVPLGLVEVK--TSNQESEV-----QEEQDPQEQTEGEGQSESEEEPE 1545

RESULT 12

T31422

C:Terminal domain-binding protein ra9 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T31422

R;Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gabara, M.; Cord

Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996

A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit

A:Reference number: 221024; MUID:96293459; PMID:8692929

A:Accession: T31422

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1473 <YUR>

A:Cross-references: UNIPROT:Q63625; EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC526

A:Experimental source: hippocampus

Query Match 3.5%; Score 129.5; DB 2; Length 1473;

Best Local Similarity 21.6%; Pred. No. 1.8;

Matches 117; Conservative 63; Mismatches 178; Indels 183; Gaps 28;

Qy 103 SRGEEEDAPTQKSLVSTAAGLFAKTRVVTSLA-----ARGGTPVAVLVLEVKLK 156

Db 238 SNGQSDADPP-----SPLSAKRVLSRSALQSHQPVAR---PVAMGLARQLPAV 284

Qy 157 AQERLDSAAPTFLEEQLMGVELLPSSLOAGLVSHRE--LDSSSSGSLSVQ--SLGNFK 212

Db 285 APEPSVEEAPVDDLGLSLCQSLWMS-SADVVIHRDGLSAKRAAPVSLQRNSVTQSR 343

Qy 213 VVSYL---LNPSYL-----DYLPOGLRCQSSAGGQVGFRTLTPESCYISED 250

Db 344 EESRLRNLQCALPSESVSGGLIGDRRPNISGLSC-----GDRALRCLPAQIV----- 392

Qy 259 GCHPQPLRAENSATARRCPP-----LSTEGLPEIHRMRWLVLFPNQGDLPDLD 311

Db 393 ---QTPVRNSDSVS-----PRSLGSLNLSDESRPKWKH-----SNSPRLN 429

Qy 312 QDN---GVHSEEEHNLRLMPQHCTONPAQAVSPAADRPETPE-KKPELVIOEVSVSPQ 367

Db 430 GSNRVVGSASTKTH-----SNFPKNIAPG--HPQKTPDRPD----- 467

Qy 368 GSSLFCELPVEKECEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGGAPSDLEASSD 427

Db 468 ----FSKLPRIKIRDGNSNSTQDQAPASQTVELPSACISRLTGRGPGQPG--RGRAD 521

Qy 428 SESEDWGEPEDDGDFDSDGSLSESDVEQSEGLHLWNSFHSVDPYKPKPNFTATITQAARI 487

Db 522 SEPSRSG--PQETGSHTSGS-----RPPAPSSH-----GNL 550

Qy 488 APRDPDSGTSMWSGSCGVSGCQEG---PIPETPDH-----SSGEEDWEPSSADEAENLKL 539

Db 551 APLGPSRGK-----GIGSSFESFRINPGNTAHCSPLSSPGFCNTFRFVDSKVQ---- 599

Qy 540 WNSFCHSEDPYLLNFKA-----PQPSGKKWKGRQDSKAS-----SEVTVP 582

Db 600 ----RKENPSPLFSIKKPKQLKSEIYDPPDTGSDSSPPSPSPESLGPDLLPSEITRTI 654

Qy 583 S 583

Db 655 S 655

RESULT 13

S38170

SRP40 protein - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YKR092c; protein YKR412a

C:Species: Saccharomyces cerevisiae

C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004

C:Accession: S38170; S40645; S37702

R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,

submitted to the Protein Sequence Database, March 1994

A:Reference number: S38158

A:Accession: S38170

A:Molecule type: DNA

A:Residues: 1-406 <BAL>

A:Cross-references: UNIPROT:P32583; EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c

A:Experimental source: strain S288C

R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J.J.

Yeast 9, 1349-1354, 1993

A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo

A:Reference number: S40644; MUID:94205265; PMID:8154186

A:Accession: S40645

A:Molecule type: DNA

A:Residues: 1-406 <BOU>

A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552

A:Experimental source: strain S288C

R;Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.

submitted to the EMBL Data Library, May 1993

A:Description: Interactions between three common subunits of yeast RNA polymerase I and

A:Reference number: S37702

A:Accession: S37702

A:Molecule type: DNA

A:Residues: 1-399,'N',401-406 <LAL>

A:Cross-references: EMBL:L11275; NID:g295670; PID:g295671

C:Genetics:

A:Gene: SGD:SRP40

A:Cross-references: SGD:S0001800; MIPS:YKR092c

A:Map position: 11R

Query Match 3.4%; Score 127.5; DB 2; Length 406;

Best Local Similarity 21.7%; Pred. No. 0.41;

Matches 74; Conservative 42; Mismatches 156; Indels 69; Gaps 10;

Qy 334 TDNPAQAVSPAADRPETPEKKPELVIOEVSVSPQSGSLFCPLPVKECEED---HTNATD 390

Db 75 SSSSSSSSSSSSSSDSESSSDSSSSSSSSSSSSSSSSSSSSSESEDETKKRARESD 134

Qy 391 LSRGESLPVSTRPVCSNKLIDYILG-----CAPSLEASSDSESDWGEPEDDG 441

Db 135 NEBAKTKKATPEPSESSSESSSESSSESSSESSSESSSSSSSSSSSSSSSESD 194

Qy 442 FDSGSLSESDVEQSEGLHLWNSFHSVDPYKPKPNFTATITQAARIAPRDPSPSGTWSG 501

Db 195 SQSSSSSSSSSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSGSSDSS 253

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 01:58:07 ; Search time 70.2452 Seconds
(without alignments)
5088.336 Million cell updates/sec

Title: US-10-650-482-4

Perfect score: 3710

Sequence: 1 METGTHRRARPGPRLGSWF.....MPNRLRIESKDLLLYSNVKK 698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3460.5	93.3	697	2	Q8BFW3 m mus muscu
2	2214	59.7	713	2	Q6P156 homo sapien
3	2213	59.6	713	2	Q96SN1 homo sapien
4	1878.5	50.6	408	2	Q6PEG0 mus musculus
5	1730.5	46.6	376	2	Q8C390 mus musculus
6	1392	37.5	407	2	Q658M2 homo sapien
7	255.5	6.9	158	2	Q6V2B6 canarypox v
8	204	5.5	657	1	M116 MOUSE
9	190.5	5.1	578	2	Q6IN02
10	189.5	5.1	578	2	Q7TQC2
11	189.5	5.1	674	2	Q75807
12	186.5	5.0	674	2	Q9NV06
13	179.5	4.8	674	2	Q6IA36
14	171.5	4.6	590	2	Q60465
15	148.5	4.0	676	2	Q9VB94 drosophila
16	147.5	4.0	725	2	Q9H986
17	147.5	4.0	872	2	Q8BQ11
18	147.5	4.0	981	2	Q8IY79
19	144.5	3.9	1444	2	Q9VTN2
20	144.5	3.9	1514	2	Q8SY55
21	144	3.9	867	2	Q13461
22	143	3.9	917	2	Q8NFN4
23	142.5	3.8	917	2	Q8WXA0
24	142	3.8	917	2	Q8IUQ1
25	142	3.8	1464	2	Q61802
26	141.5	3.8	1546	2	Q7Z313
27	141	3.8	522	2	Q59740
28	141	3.8	1198	2	Q5ZRM5
29	140.5	3.8	822	1	MCPI1 MOUSE
30	140.5	3.8	967	2	Q920Q9
31	139.5	3.8	990	2	Q811R2

ALIGNMENTS

32	138	3.7	1455	2	Q62PU4	Q62pu4 mus musculu
33	138	3.7	1578	2	Q6I6G8	Q6i6g8 mus musculu
34	138	3.7	1581	2	Q9P2P5	Q9p2p5 homo sapien
35	137.5	3.7	1093	2	Q8NFN5	Q8nfn5 homo sapien
36	136.5	3.7	1411	1	Y329_HUMAN	O15040 homo sapien
37	136.5	3.7	3925	1	BSN_HUMAN	Q9UP65 homo sapien
38	135	3.6	998	2	O16643	O16643 caenorhabdi
39	134	3.6	754	2	Q6PEI3	O6pei3 brachydanio
40	133.5	3.6	368	2	Q6CNF1	O6cnf1 kluyveromyc
41	133.5	3.6	782	2	O16883	O16883 caenorhabdi
42	133.5	3.6	883	1	PCBR_RAT	P55068 rattus norv
43	133.5	3.6	1445	2	Q6I6F5	O6i6f5 mus musculu
44	133	3.6	1586	2	Q6NZN8	Q6nzn8 mus musculu
45	132.5	3.6	682	2	Q98PI8	Q98pi8 mycoplasma
46	132.5	3.6	1095	2	Q9HCG6	Q9hcg6 homo sapien
47	132.5	3.6	1784	2	Q8WML3	Q8wml3 macaca fasc
48	132	3.6	1333	2	Q76C96	Q76c96 bacillus cl
49	132	3.6	2259	2	Q7XIF9	Q7xif9 oryza sativ
50	131.5	3.5	509	2	Q67080	Q67q80 symbiobacte
51	131.5	3.5	1459	2	Q86VP9	Q86vp9 homo sapien
52	131.5	3.5	1600	2	Q6NZJ6	Q6nztj6 mus musculu
53	131.5	3.5	2055	2	Q9Y4G4	Q9y4g4 homo sapien
54	131.5	3.5	4169	2	Q8TCU4	Q8tcu4 homo sapien
55	131	3.5	1049	2	Q9ZNR9	Q9znz9 volvox cart
56	131	3.5	1808	2	Q7SDM2	Q7sdm2 neurospora
57	130	3.5	613	2	Q9UEB1	Q9ufb1 homo sapien
58	130	3.5	1560	2	Q26644	Q26644 strongyloce
59	129.5	3.5	1473	2	Q63625	Q63625 rattus norv
60	129	3.5	930	2	Q925G9	Q925g9 mus musculu
61	129	3.5	1199	2	Q8F4X3	Q8f4x3 leptospira
62	129	3.5	1216	2	Q68DF8	Q68df8 homo sapien
63	129	3.5	1229	2	Q68H99	Q68h99 sus scrofa
64	129	3.5	1503	2	Q8IWU2	Q8iwt2 homo sapien
65	128.5	3.5	1051	2	Q80TA3	Q80ta3 mus musculu
66	128.5	3.5	1404	2	Q80VM7	Q80vm7 mus musculu
67	128	3.5	908	2	Q6CST7	Q6cst7 kluyveromyc
68	128	3.5	1199	2	Q72QW8	Q72qw8 leptospira
69	127.5	3.4	406	1	SR40_YEAR	P32583 saccharomyc
70	127.5	3.4	568	2	Q7T3E4	Q7t3e4 brachydanio
71	127.5	3.4	1003	2	Q948S6	Q948s6 nicotiana t
72	127.5	3.4	1078	2	Q6ZTV6	Q6ztv6 homo sapien
73	127.5	3.4	1281	2	Q6ZRR1	Q6zrr1 homo sapien
74	127	3.4	919	2	Q6ZPY9	Q6zpy9 mus musculu
75	127	3.4	920	2	Q9P232	Q9p232 homo sapien
76	127	3.4	996	2	Q86YN3	Q86yn3 homo sapien
77	127	3.4	1002	2	Q86YN4	Q86yn4 homo sapien
78	127	3.4	1017	2	Q86YN5	Q86yn5 homo sapien
79	127	3.4	1023	2	Q86YN6	Q86yn6 homo sapien
80	127	3.4	1056	2	Q95621	Q95621 homo sapien
81	127	3.4	1476	2	Q9UP83	Q9up83 homo sapien
82	126.5	3.4	371	2	Q6FTH7	Q6fth7 candida gla
83	126.5	3.4	1841	2	Q6NZK2	Q6nzk2 mus musculu
84	126.5	3.4	2272	2	Q74ZJ8	Q74zj8 ashbya goss
85	126	3.4	365	2	Q8C4A1	Q8c4a1 mus musculu
86	126	3.4	817	1	NEB2_RAT	Q35274 rattus norv
87	126	3.4	817	2	Q6R891	Q6r891 mus musculu
88	126	3.4	871	2	Q6COL3	Q6col3 kluyveromyc
89	126	3.4	1213	2	Q7TT28	Q7tt28 mus musculu
90	126	3.4	1229	2	Q9SMV0	Q9smv0 arabidopsis
91	126	3.4	1306	2	Q7TT18	Q7tt18 mus musculu
92	126	3.4	1317	2	Q91Y25	Q91y25 cricetus
93	126	3.4	1494	2	Q756B4	Q756b4 ashbya goss
94	126	3.4	1726	2	Q6ZPP9	Q6zpp9 mus musculu
95	126	3.4	2764	2	Q7Z7D6	Q7z7d6 homo sapien
96	125.5	3.4	774	2	Q6Y685	Q6y685 mus musculu
97	125.5	3.4	1100	2	Q6ZPF9	Q6zpf9 mus musculu
98	125.5	3.4	1158	2	Q97169	Q97169 drosophila
99	125.5	3.4	1172	2	Q6P9B4	Q6p9b4 homo sapien
100	125.5	3.4	1198	1	UB42_HUMAN	Q9h9j4 homo sapien

RESULT 1
Q8BFW3 PRELIMINARY; PRT; 697 AA.
AC Q8BFW3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length enriched
DE library, clone: C33093D15 product:hypothetical protein, full insert
DE length enriched library, clone: C53002L24 product:hypothetical
DE protein, full insert sequence).
GN Name=Pppl15b;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata K., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao K., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura K., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
[5]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Spinal cord;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao K., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura K., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
[6]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,

RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[7]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal cord;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK049028; BAC33517.1; -;
DR EMBL; AK082957; BAC38708.1; -;
DR MGI; MGI:2444211; Pppl15b.
DR GO; GO:000164; C:protein phosphatase type 1 complex; IDA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0045182; F:translation regulator activity; IC.
DR GO; GO:0006983; P:BR-overload response; IDA.
DR GO; GO:0006446; P:regulation of translational initiation; IC.
DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.
DR GO; GO:0006979; P:response to oxidative stress; IDA.
KW Hypothetical protein.
SQ
SEQUENCE 697 AA; 77711 MW; E439B12615F3737 CRC64;

Query Match 93.3%; Score 3460.5; DB 2; Length 697;
Best Local Similarity 94.3%; Pred. No. 8e-204;
Matches 658; Conservative 7; Mismatches 32; Indels 1; Gaps 1;
Qy 1 METGTHARKRPGRLGSGWFLPFLRRSHACSFPPPPSSRQNGNSALPERRTRYWTKL 60
Db 1 METGTHARKRPGRLGSGWFLPFLRRSHACSFPPPPSSRQNGNSALPERRTRYWTKL 60
Qy 61 LSQALLLPSPFQKLLWSQLSGGLIPTRWLDFAAYSALRASGRRESDAPTVQKLSY 120
Db 61 LSQALLLPSPFQKLLWSQLSGGLIPTRWLDFAAYSALRASGRRESDAPTVQKLSY 120
Qy 121 TAAGLFAKTRVSTLALARGSTPVAVLVRLVKLKAQERALDSAAFTLEEQQLWGV 180
Db 121 LRLD-SEEDLVVSSDLWLEGLQWQCSSDLELKLKAQERALDSAAFTLEEQQLWGV 179
Qy 181 LPSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQSSAGG 240
Db 180 LPSLQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCQSSAGG 239
Qy 241 QFVGFRITLPESCYSLEDGCHPOPLRAEMGATARRCPPLSTEGLPRIHRRRWLVFLQ 300
Db 240 QFVGFRITLPESCYSLEDGCHPOPLRAEMGATARRCPPLSTEGLPRIHRRRWLVFLQ 299
Qy 301 PNOGQDLPDQNGVHSLFEEHNLRLMDQHCNTDPAQAVSPAADRPETPEKKPELVQ 360
Db 300 ANKQGEIPLTDQNGVHSLFEEHNLRLMDQHCNTDPAQAVSPAADRPETPEKKPELVQ 359
Qy 361 EVSQSQGSSLFCELPVKECEEDHTNATDLSDRGESLPVSTRPVCNKLDIYILGAPS 420
Db 360 EVSQSQGSSLFCELPVKECEEDHTNATDLSDRGESLPVSTRPVCNKLDIYILGAPS 419
Qy 421 DLEASSDSESDWGEEDDGFSDGSLSDVDQSEGLHLWNSFHSVPDPKQNTAT 480
Db 420 DLEASSDSESDWGEEDDGFSDGSLSDVDQSEGLHLWNSFHSVPDPKQNTAT 479
Qy 481 IQTAARTAPDPSDGTSGVSGCQSGCPGPETPDHSGSGEDDWEPSDAENLKLW 540
Db 480 IQTAARTAPDPSDGTSGVSGCQSGCPGPETPDHSGSGEDDWEPSDAENLKLW 539


```

Qy 541 NSPCHSDPYNLLNFKAPFOPSGKNWGRQDSKASSEVTVAFSGHHTLLSKCKQLLESQE 600
Db 540 NSPCHSDPYNLLNFKAPFOPSGKNWGRQDSKASSEATVAFSGHHTLLSKCKQLLESQE 599
Qy 601 DNCPCGGLGALAGERYTHIKRKKVTFLEEVTEYIISGDDRKGPWEFFARDGCRFOKRI 660
Db 600 DNCPCGGLGALAGERYTHIKRKKVTFLEEVTEYIISGDDRKGPWEFFARDGCRFOKRI 659
Qy 661 QETEVAGYCLAFEHREKMFNRLIESKDLLLYSNVKK 698
Db 660 QETEVAGYCLAFEHREKMFNRLIESKDLLLYSNVKK 697

RESULT 2
Q6P156 PRELIMINARY; PRT; 713 AA.
ID Q6P156 AC Q6P156
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Protein phosphatase 1, regulatory subunit 15B.
GN Name=PPP1R15B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUR=Eye;
RC MEDLINE=24238257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUR=Eye;
RA Director MGC Project;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065280; AAB65280.1;
SQ SEQUENCE 713 AA; 79125 MW; 76C2DA38F9E76A85 CRC64;

Query Match 59.7%; Score 2214; DB 2; Length 713;
Best Local Similarity 64.8%; Pred. No. 2.2e-127;
Matches 454; Conservative 55; Mismatches 174; Indels 18; Gaps 7;

Qy 1 METGTHARKPGRPLGSRFLPPL-RRSHACSEFPFPPSSRONPGN-----SALPERRT 54
Db 1 NEPTGGSRKKLGPAGRFPPFPFPPRSQAGSKFPPTPLGPENSGNPTLLSSAQPETRV 60
Qy 55 RYWTKLLSQLALAPSLPQKLLNSQLSGGLIPTRWLDFAASYGALRASRGRESDAPTV 114
Db 61 SYWTKLLSQLALAPLGLLQKLVNSQLFGGFMFTRWLDFAVYSGALRALKGREKPAAPTA 120
Qy 115 QKSLSYTAAGLFAKTRVVSTLALAGGTPVAVLVRLEVKLKAERALDSAPTFLEEQ 174
Db 121 QKSLSSQLD-SSDPVSPTSPLDWLEEGTHWQYSPDPLKLELKAGSALDPAAPAFLEEQ 179

```

```

Qy 175 LMGVELLPSSLAQGLVSHRELSSSGPLSVQSLGNFKVSVYLLNPSYLDYLPQLGLRCQ 234
Db 180 LMGVELLPSSLAQGLVSHRELSSSGPLSVQSLGNFKVSVYLLNPSYLDYLPQLGLRCQ 239
Qy 235 SSAGGQGVFPRLTTPESCYLSDGCHPPLRAEMSATARRCPPLSTGLPIHHRMR 294
Db 240 NSDGNSEVVGFQTLTPESCLREDHCHPPLSABLIPASWGCPPLSTGLPIHHRMR 299
Qy 295 WLVL-QPNQGDLPDLQDNGYHSLSEEHNLRLMDPQHCTDNPQAQVPAADRP-----E 349
Db 300 RLEFLQASKGDLPTPDQDNGYHSLSEEHNLRLMDPKHCRDNPQVPAAGDIPGNTQE 359
Qy 350 PTEKKPELVIEV-----SQSPGSSLPFCELPVEKECEDHTNATLSDRGESLPVSTRP 404
Db 360 STEEKIELLTTEVPLALEBESSEGCPSSEIPMEKEFGEGRISVVSVYSLYEGDLPIARP 419
Qy 405 VCSNKLIDYILGGAPSDLEASSDSESDGEBEDDGFSDGSLSDSDVEQDSGLHLWN 464
Db 420 ACSNKLIDYILGGASSDLETSDEGEDWDEAEDDGFSDGSLSDSDLEQDPEGLHLWN 479
Qy 465 SFHSVDYPKQNPFTATQTAARIAPRDPDSGTSWSGCGV-GSCQEGPLPETPDHSSGE 523
Db 480 SFCSVDYPNPQNPFTATQTAARIVPEPSDSEKDLGSKDLENSQSGSLPETPEHSSGE 539
Qy 524 EDDWPSADAENLKLWNSFCHSEDYPNLLNFKAPFOPSGKNWGRQDSKASSEVTVAFS 583
Db 540 EDDWESSADAESLKLWNSFCHSEDYPNLLNFKAPFOPSGKNWGRQDSKASSEVTVAFS 599
Qy 584 GHHTLLSCKAQLLESQEDNCPGGLGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLUSCKVQLLGSESECPDSVQDVLSSGGRRHTVKKRKKVTFLEEVTEYIISGDEDRK 659
Qy 644 GPWEFFARDGCRPKKIQETEVAGYCLAFEHREKMFNRLR 684
Db 660 GPWEFFARDGCRPKKIQETEVAGYCLAFEHREKMFNRLQ 700

RESULT 3
Q96SN1 PRELIMINARY; PRT; 713 AA.
ID Q96SN1 AC Q96SN1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Hypothetical protein FLJ14744.
DB Hypothetical protein FLJ14744.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Oka T., Suzuki Y., Nishikawa T., Otauki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayaashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishi S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Inose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,

```

RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs".
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK027650; BAB5266.1; --
DR Genew; HGNC:14951; PPRI15B.
SQ SEQUENCE 713 AA; 79125 MW; 26C2D06144AAD25E CRC64;

Query Match 59.6%; Score 2213; DB 2; Length 713;
Best Local Similarity 64.8%; Pred. No. 2.6e-127; Indels 18; Gaps 7;
Matches 454; Conservative 55; Mismatches 174;

QY 1 METGTHARKRPGRLGFWFLPFL-RRSHACSSFFPPSSRQNPNGN-----SALPERRT 54
Db 1 MEPGTGSRKRLGPRAGRFPPFPFRPSQAGSSKFTPLGPENSGNFTLSSAQPETRV 60

QY 55 RYWKLLSLLALPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESDAPTV 114
Db 61 SYWTKLLSLLAPLPGLLQKVLWSQLSGGLIPTRWLDFAAGYSALRALKGRKPAAPTA 120

QY 115 OKSLSYTAAGLFAKTRVVTIALARGCTPVAVLVLRLEVKLKAQERLDAAPFLLEQQ 174
Db 121 QKSLSLLQD--SSDPSTVSPDLWEEGHHQVSPDPLKELKAGSALDPAQAFLLLEQQ 179

QY 175 LWGVLLPSSLQAGLVSHRELDSSGSLVSQSLGNFKVSVYLLNPSYLDLPOLGLRCQ 234
Db 180 LWGVLLPSSLQSLRYSNRELSSGSLVSQSLGNFKVSVYLLNPSYLDLPOLGLRVSQ 239

QY 235 SSAGGGQVGRITLTPESCYLSGDCGCHPQPLRAEMSAATWRCPPPLSTEGPLPEIHRMR 294
Db 240 NSDGNSEVVGFTLTPESCYLSGDCGCHPQPLRAEMSAATWRCPPPLSTEGPLPEIHRMR 299

QY 295 WLVL- QPNQODLPTLDQNGYHSLREHNLMDPOHCTDNPQAQVSPAADRP---E 349
Db 300 RLEFLQQAASKGQDLPPTPDQNGYHSLREHNLMDPOHCTDNPQAQVSPAADRP---E 359

QY 350 PTEKPELVIQEV-----SQSPQSSSLFCPLPVEKECEEDHTNADLSDRGESLPVSTRP 404
Db 360 STEEKIELLTVEVPLALEEFSPECCPSSEIPEMEKEGGRISVVDYVYLSGDLPIASRP 419

QY 405 VCSNKLIDYILGGAPSDLEASDSSEDPWGEDDGDGSLSESVDQSDSGLHLWN 464
Db 420 ACSNKLIDYILGGASSDLETSPDEGDWDEAEDGDFSDSLSDSLEQDPEGLHLWN 479

QY 465 SFHSVDPPKPNFTATIQTAARIAPRDSGTSWSGCGV-GSCQEGPLPETDHSSE 523
Db 480 SFCSDVPNPNFTATIQTAARIPEEPSDSSEKOLSGKSDLENSQSGLSPETPEHSSGE 539

QY 524 EDDNEPSADENLKLWNSFCHSEDPYNLLNFKAPQPSGKNWKGRODSKASSEVTVAFS 583
Db 540 EDDWESSADEASLKLWNSFCNSDDPNPLNFKAPQPSGKNWKGRODSKASSEVTVAFS 599

QY 584 GHHTLLSCAQLLESQBNCPGCGLGEALAGERYTHIKRKKVTFLEEVTEYIISGDEDRK 643
Db 600 ECHTLLSCVKQLLGSQSECPDQVDRDVLSSGGRHTRVKKVTFLEEVTEYIISGDEDRK 659

QY 644 GPWEFARDGCRFQKRIQETVATGYCLAFEHREKMNRLR 684
Db 660 GPWEFARDGCRFQKRIQETDALTGYCLTFEHRERMFNRLQ 700

RESULT 4

Q6PEGO

ID Q6PEGO PRELIMINARY; PRT; 408 AA.

AC Q6PEGO;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE PP1r15b protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; Pubmed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX Strausberg R.;
RL Submitted (Sep-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC058078; AAH58078.1; --
DR GO; GO:000164; C:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0045182; F:translation regulator activity; IC.
DR GO; GO:0006983; P:ER-overload response; IDA.
DR GO; GO:0006446; P:regulation of translational initiation; IC.
DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.
DR GO; GO:0006979; P:response to oxidative stress; IDA.
SQ SEQUENCE 408 AA; 45432 MW; 44AA70EFE25796C4 CRC64;

Query Match 50.6%; Score 1878.5; DB 2; Length 408;
Best Local Similarity 90.6%; Pred. No. 4e-107;
Matches 367; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

QY 1 METGTHARKRPGRLGFWFLPFLRRSHACSSFFPPSSRQNPNGNSALPERRTRYWTKL 60
Db 1 METGTHARKRPGRLGFWFLPFLRRSHACSSFFPPSSRQNPNGNSALPERRTRYWTKL 60

QY 61 LSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESDAPTVOKSLSY 120
Db 61 LSQALLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESDAPTVOKSLSY 120

QY 121 TAGLFAKTRVVTIALARGCTPVAVLVLRLEVKLKAQERLDAAPFLLEQQWGVLE 180
Db 121 LRLD-SSEDLVVSSLDWLEEGQLQWQSSSDLELKLKAQERLDAAPFLLEQQWGVLE 179

QY 181 LPSSLQAGLVSHRELDSSGSLVSQSLGNFKVSVYLLNPSYLDLPOLGLRCOSSAGG 240
Db 180 LPSSLQAGLVSHRELDSSGSLVSQSLGNFKVSVYLLNPSYLDLPOLGLRCOSSAGG 239

QY 241 QFVGFRITLTPESCYLSGDCGCHPQPLRAEMSAATWRCPPPLSTEGPLPEIHRMRWLVFLQ 300
Db 240 QFVGFRITLTPESCYLSGDCGCHPQPLRAEMSAATWRCPPPLSTEGPLPEIHRMRWLVFLQ 299

QY 301 PNOQODLPTLDQNGYHSLREHNLMDPOHCTDNPQAQVSPAADRPETEKPELVIQ 360
Db 300 ANKQELPTPDQNGYHSLREHNLMDPOHCTDNPQAQVSPAADRPETEKPELVIQ 359

```

QY 361 EVSQSQGSLFCBLPVEKECEDHTWTATDLSRGSPLVSTRPV 405
DB 360 EVSQSQGSLFCBLPVEKECEDHTWTATDLSRGSPLVSTRPV 404

RESULT 5
Q8C390 PRELIMINARY; PRT; 376 AA.
AC Q8C390;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:D930040107 product:hypothetical protein, full insert
DE sequence. (Fragment).
GN Name=Pp1r15b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitauai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto K., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hangsaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,

```

```

RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK086506; BAC39701.1; -.
DR MGD; MGI:2444211; Pp1r15b.
DR GO; GO:0000164; C:protein phosphatase type 1 complex; IDA.
DR GO; GO:0004722; P:protein serine/threonine phosphatase activity; IDA.
DR GO; GO:0045182; P:translation regulator activity; IC.
DR GO; GO:0006983; P:ER-overload response; IDA.
DR GO; GO:0006446; P:regulation of translational initiation; IC.
DR GO; GO:0042542; P:response to hydrogen peroxide; IMP.
DR GO; GO:0006979; P:response to oxidative stress; IDA.
KW Hypothetical protein.
FT NON_TER 376
SQ SEQUENCE 376 AA; 41949 MW; 972525D3D472C8F8 CRC64;
Query Match 46.6%; Score 1730.5; DB 2; Length 376;
Best Local Similarity 89.9%; Pred. No. 4.3e-98;
Matches 339; Conservative 7; Mismatches 30; Indels 1; Gaps 1;
QY 1 METGTHRAKRPGPRLGSWFRPFLRRSHACSSSEFPSPSRQNGNSALPERTRYWTKL 60
DB 1 METGTHRAKRPGPRLGSWFRPFLRRSHACSSSEFPSPSRQNGNSALPERTRYWTKL 60
QY 61 LSOLLALLPSLFQKLLWSQLSGGLIPTRWLPDPAASYSALRASRGRESDAPTVOKSLSY 120
DB 61 LSOLLALLPSLFQKLLWSQLSGGLIPTRWLPDPAASYSALRASRGRESDAPTVOKSLSY 120
QY 121 TAAGLFAKTRVVSITLALARGGTVAVLRLVKLKAQERALDSAAPTFLLEQLMGVEL 180
DB 121 LRLD-SSEDLVSSLDWLEGLQWCSSSDLELKLKAQERALDSAAPTFLLEQLMGVEL 179
QY 181 LPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCOSSAGG 240
DB 180 LPSSLOAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYLDYLPQLGLRCOSSAGG 239
QY 241 QFVGFTLTPESCYLSGDCGCHPQPLRAEMSAATARRCCPLSTGLPEIHHRMRWLVFLQ 300
DB 240 QFVGFTLTPESCYLSGDCGCHPQPLRAEMSAATARRCCPLSTGLPEIHHRMRWLVFLQ 299
QY 301 PNOGQDLPTLDQDNGYHSLSEEHNLRLMDPQHCTDNPAAQVSPAADRPETEKKPVLVIQ 360
DB 300 ANKGQELPTDQDNGYHSLSEEHNLRLMDPQHCTDNPAAQVSPAADRPETEKKPVLVIQ 359
QY 361 EVSQSQGSLFCBLPVEKECEDHTWTATDLSRGSPLVSTRPV 377
DB 360 EVSQSQGSLFCBLPVEKECEDHTWTATDLSRGSPLVSTRPV 376

RESULT 6
Q658M2 PRELIMINARY; PRT; 407 AA.
AC Q658M2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp666i1186 (Fragment).
GN Name=DKFZp666i1186;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stomach;
RG The German cDNA Consortium;
RA Bahr A., Lauber J., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA Han M., Wiemann S.;

```



```

QY 280 LSTEGLEPEIHHRRMRWLVLFPQNGQDLPTLDQNGHSLSEENHNLMDPQHCTDNPQAQ 339
DB 200 TPVPFLGEAEHQATE-----EKGTENKADPNPSGSGHSRAWEY-YGREKPKQGEAKVE 254
QY 340 AVSPAADRP-----EPTKKPELVIQEVSQSQSGSLFCELPVKECEEDHNTATDLSDRG 395
DB 255 AHRAGQGHPCRNAEAEAGGPTTF--VCTGNFLKAWVYRPGEDTEEDSDSDSAEDT 312
QY 396 ESULPVSTRPVCSNKLIDYILGGAPSLDEASDSSESD-----W-----GE 435
DB 313 AQTGATPHTSAFLKAWVYRPGEDTEEDSDSDSAEDTAQTGATPHTSAFLKAWVYRPG 372
QY 436 EPEDDGDSDGLSSESDVEQDSEGLHLWNSFHSVDYKPNQFTATITQTAARIAPRDPDS 495
DB 373 DTEENSDDLSS--ABEDTAQTGATPHT--SAFLKAWVYRPGEDTEEDSDSDSAEDTAQT 429
QY 496 GTWSGSCGVGSCQEGPLPTPDHSSGEEDDWEPSADEAENLKLWNSFCHSEDPYNLNF 555
DB 430 GATPHTSPFLKAWVYRPGEDTEEDSDSDSAEDTADSSQSPC----- 478
QY 556 KAPQPS-----GKNWKGRQDSKASSEVTVAFSGHH-----TLLSCKAQLES-- 598
DB 479 --LQPCRLPGKTKGRGEPPLFQVAFVLPGEKPSPWAAPKLPLRLQRLRLFKAPT 535
QY 599 --QEDNCPGCGLGALAGERYTHIKKVKVTFLEVTYIIS-----GDEDRKGWBEFAR 651
DB 536 RDQDPEIP-----LKKARKVHFAEKVTVHFLAVWAGPAQAARRGPWEQFAR 580
QY 652 DGCRCFKRKIOBTEVAIGVCLAFEHREKMFNRLR 684
DB 581 DRSRFRARIAQAEKUGPYLTPDSSRARAWRLR 613

RESULT 9
Q6IN02 PRELIMINARY; PRT; 578 AA.
AC Q6IN02
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Myd116 protein.
GN Name=Myd116;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC TISSUS=Heart;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN ONCOGENE 22:3827-3832(2003).
RC SEQUENCE FROM N.A.
TISSUS=Heart;

```

```

RA Strauberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072513; AAT72513.1; -.
SQ SEQUENCE 578 AA; 63583 MW; D9387CA71923345F CRC64;

Query Match 5.1%; Score 190.5; DB 2; Length 578;
Best Local Similarity 22.8%; Pred. No. 0.002;
Matches 131; Conservative 58; Mismatches 214; Indels 171; Gaps 30;

QY 216 YLLNPSYLDLPQLGLR-----COSSAGGQGVGFRTLTTPESCYSLSBDGCHPQP 264
DB 20 YLLSP-LMGFLSRWSRLRGPEVSEAWLAETVAGANQIEADALLTTPP--VSENL---P 73
QY 265 LR-----AEMSATAWRR-----PPLSTGLPEIHHRRMRWLVLFPQNGQDLPTL 310
DB 74 LRETEGNGTPWESKAAQRLCLDVEAQSSPKTKWGLSDIDHNGK-----PQD----- 121
QY 311 DDNGYHSLSEEH--NLLRMDPOHC--TONPAQAV-----SPAADR 347
DB 122 ----GLREQVEHTAGLPTLQPLHQQADKKGVEGVAREEGVSELAYPTSHWEGGPAEDE 177
QY 348 PEPTKKPELVIQEVSQSP---QGSSILFCELPVKECEED-----HTNATDLS 392
DB 178 EDTETVKKAQAASASIAQYKPESTSVYCPGEAEHRAATEKGTDNKAEPGSGHSRVWEYH 237
QY 393 DR-----GESLPVSTRP---VCSNKLIDYILGGAPSLDEASDSSESDW-----GEEPE 438
DB 238 TREPRKQEGETKEQHRAGQSHPCNAEAE---EGGPETSVCSGSAFLKAWVYRPGEDTE 294
QY 439 DDGFDSDGSLSESDVEQDSEGLHLWNSFHSVDYKPNQFTATITQTAARIAPRDPDSGTS 498
DB 295 EEE--DSDLDSEAEDTAHTCTTPT--SAFLKAWVYRPGEDT-----BEEDDG-D 339
QY 499 WSGSCGVGSCQEGPLPET-----PDHSGEEDDWEPSAD-EAENLKLWNSFCH-- 545
DB 340 WD-SAEEDTAQSCCTTHTSAFLKAWVYRPGEDTEEDSDSENAPVSETVDSQSTQCL 398
QY 546 -----SEDPYNLLNFKAPFPQSGKNWKGQDSKASSEVTVAFSGHHTLLSCKAQL 595
DB 399 PVKTKGCGEAEPPFQVAFYLPQKPAPEW-----AAPKLPLRL--QKRLRSFKAP- 448
QY 596 LESQDNCPCGGLGEALAGERYTHIKKVKVTFLEVTYIIS-----GDEDRKGWBEFA 650
DB 449 ARNQDPEIP-----LKKRKVHFSEKVTVHFLAVWAGPAQAARRGPWEQFA 493
QY 651 RDCRCFKRKIOBTEVAIGVCLAFEHREKMFNRLR 684
DB 494 RDRSRFRARIAQAEKUGPYLTPAPARAWRLR 527

RESULT 10
Q7TQC2 PRELIMINARY; PRT; 578 AA.
AC Q7TQC2
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Protein phosphatase regulatory subunit 15A.
GN Name=Gadd34;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=22697643; PubMed=12813455; DOI=10.1038/sj.onc.1206567;
RA Hollander M.C., Poole-Kella S., Fornace A.J. Jr.;
RT "Gadd34 functional domains involved in growth suppression and
RT apoptosis."
RL Oncogene 22:3827-3832(2003).
DR EMBL; AY128642; AAM77795.1; -.
SQ SEQUENCE 578 AA; 63569 MW; 923BC49921C0BC61 CRC64;

```

```

Query Match          5.1%; Score 189.5; DB 2; Length 578;
Best Local Similarity 22.8%; Pred. No. 0.0023;
Matches 131; Conservative 58; Mismatches 214; Indels 171; Gaps 30;

QY 216 YLLNPSYLDYLPQLGLR-----CQSSAGGQGVFGRTLPESCYLESDGCHPQP 264
D 20 YLLSP-LMGFLSRAWSRLRGPVSEAWLAETVAGANQIEADALLTPPP--VSENLH---P 73

QY 265 LR-----AENSATAWRRC-----PPLSTEGLEPIHHRMRWLVLQPNQODLPTL 310
D 74 LRETEGNTPEWSKAAQRLCLDVEAQSPPKTWGLSLDIEHNGK-----PGQD---- 121

QY 311 DDQNGYHSLERH--NLLRMDPHOC--TDNPAQAV-----SPAADR 347
D 122 ---GLRQEVHTAGLTQLQLHLOGADKKVGEVVARBEGVSELAYPTSHWEGCPAED 177

QY 348 PEPTKKPVLVIQVQSP--QSSSLFCPLPVEKCEED-----HTNATDLS 392
D 178 EDTETVKKAHQAASAASIAPGVKPSVSVCPGEAHRATEEKGTONKABPSSGSHSRVWEYH 237

QY 393 DR-----GESLPVSTRP-----VCSNKLDYILGAPSDLEASDSESDW----GEEPE 438
D 238 TRBRPKQGETKPEQHRAGQSHPCQNAE--EGGPETSVCSGSAFLKAWVYRPGEDTE 294

QY 439 DDGFSDCSLGESDVEQDSEGLHLWNSPHSVDPYKPNFTATQTAARIAPRDPDSGTS 498
D 295 EEE-DSLDLSAEDTAHTCTTPHT-SAPLKAWVYRPGEDT-----EEDDG-D 339

QY 499 WSGCGVSGCQGLPET-----PDHSSGEEDWEPSPAD-EAENKLMNSFCH-- 545
D 340 WD-SAEEDASQCTTPHTSAFLKAWVYRPGEDTEEDDSENVAPVDSETVDSQCSTQHCL 398

QY 546 -----SEDPNLNFAPQPSGKNWKGQDSKASSEVTVAFSGHHTLLSCKAQL 595
D 399 PVEKTKGCEAEPPFPQVAFYLPQCKAPPW-----AAPKLPLRL--QKRLSRFKAP- 448

QY 596 LESQEDNCPGCGLGEALAGERYTHIKKKVTFLEEVTEYIS-----GDEBKGPWEFA 650
D 449 ARNQDPEP-----LKGKRVHSEKVTYHFLAVWAGPAQAARGFWEQFA 493

QY 651 RDGCRFQKRIQETVAIGCYCLAFHREKMFNRLR 684
D 494 RDRSRFARRIAQABEQGLPYLTAPFARAWTLRL 527

RESULT 11
O75807 PRELIMINARY; PRT; 674 AA.
AC O75807;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Apoptosis associated protein (protein phosphatase 1, regulatory subunit 15A).
GN Name=GADD34; Synonyms=PPP1R15A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97298078; PubMed=9153226; DOI=10.1074/jbc.272.21.13731;
RA Hollander M.C., Zhan Q., Bae I., Fornace A.J. Jr.;
RT "Mammalian GADD34, an apoptosis- and DNA damage-inducible gene.";
RL J. Biol. Chem. 272:13731-13737(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Iaquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalko U., Schmutz D.E., Schnerch A., Schein J.E.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RC SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83981; AAC25633.1; -.
DR EMBL; BC003067; AAH03067.1; -.
DR Genew; HGNC:14375; PPIR15A.
DR GO; GO:0006915; P:apoptosis; TAS.
DR GO; GO:0007050; P:cell cycle arrest; TAS.
DR GO; GO:0006974; P:response to DNA damage stimulus; TAS.
SQ SEQUENCE 674 AA; 73477 MW; B257AAL1456D1403 CRC64;

Query Match          5.1%; Score 189.5; DB 2; Length 674;
Best Local Similarity 21.5%; Pred. No. 0.0028;
Matches 156; Conservative 74; Mismatches 269; Indels 227; Gaps 32;

QY 60 LLSQLALLPSLFQKLLNLSQLSGGLIPTR-WLDFPAAYSALRASRGREESDAPTVOKSL 118
D 21 LLSFVPMGLLR-----ANSRLR-GLGLEPFWLVEAVKGAAL----- 55

QY 119 SYTAAGLFAKTRVVTTLARGGTPVAVLRLVLEVKLKAQERALDSAAP-----T 168
D 56 --VEAGLEGAR-----TPLA--IPHTFWRPRPEEAEDSGPGEDRETGLKT 100

QY 169 FLLEQQLWGVLLPSSLAQAGLVSHRELDSSSGPLSVQSLGNFKVSYLLNPSYL----- 223
D 101 SSSLPEAWG--LLDD--DDGMGEREATSVPRGQSQFADGQ-----RAPLSPLLIRTLQ 152

QY 224 --DYLPGQLGRCSAGGQGVGFRTLT-----PESCYLESDGCHPQPLAEMSAATA--- 273
D 153 GSDKNP--GEEKAEAGEVAGEEGVKNFSPYPPSHRECCPAVEEEDDEEAVKKBAHRTSTA 210

QY 274 -----WRRCP-----PLSTEGLEPIHHRMRWLVLQPNQODLPTLDQNGYHSL 320
D 211 LSPGSKPSTWVSCPGEENQATDKRTSKGARKTSVSPRSGSDPSRSEYRSGEASEE 270

QY 321 EHNLLRMDPHOCHTDNPAQAVSPAADRPEPEKPELVIQEVQSQSSGLFCPLPVEKE 380
D 271 KE-----EKAKHETGKGEAA-----PGQSSAP-----AQRPLKQWVCQPSDEEE 311

QY 381 CERDHTNATDLSRGESLPVSTRPVCSNKLDYILGAPSDLEASDSESDWGEEREDD 440
D 312 GEVKALGAEEKGAEACPCPCIPPSAPFLKAWVWPG---EDTBEEDDEE---DESDS 364

QY 441 GFDSDGSLSDVEQDSEGLHL--WNSFHSVDPYKPNFTATIQ-----TAARIAPRDP 493
D 365 GSDEEAGEAEASSTPATGVFLKSW-----VYQGEDTEDEDESDSGSADEDEAE 417

QY 494 DSGTSMGSCGVSGCQGLPPTPDH-----SGSEDDWEPSEADEAENKLMNSFCHSE 547
D 418 TSASTPPASAFKAWVYRPGEDTEDEDEDDVDSEKDDSEAAALGEARS----- 466

QY 548 DPNYLLNFKAPFOPSG-----KNWKGQDSKASSEVTVAFSGHHTLLSCKAQLLESQDN 603
D 467 DPH-----PSHPDQRAHFRGWYRPGKETEE-----EAEAD--- 498

```

```

QY 604 PGCGLGEA-----LAGER-----YTHIKRK 623
DB 499 -----WGEAECPFRVAIVYFGEKPPPPWAPRLRLQRLKRPETHTDPDPTPKAR 554
QY 624 KVTFLBVTYYIS-----GDEDRKGWBEFARGCFKRIQIETVAIGYCLAFEHREK 678
DB 555 KVFSEKVTYHFLVWAGPAQAARQGPWEQLARDRSFARRITQAOBELSPCLTPAAR 614
QY 679 MFNRLR 684
DB 615 AWARLR 620

RESULT 12
Q9NVU6 PRELIMINARY; PRT; 674 AA.
ID Q9NVU6
AC Q9NVU6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10499.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Iehibaishi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano Y.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Teraashina Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagae T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK001361; BAA91649.1; -.
SQ SEQUENCE 674 AA; 73363 MW; 8F9808E873B29CBFA CRC64;

Query Match 5.0%; Score 186.5; DB 2; Length 674;
Best Local Similarity 21.5%; Pred. No. 0.0043;
Matches 156; Conservative 73; Mismatches 270; Indels 227; Gaps 32;

QY 60 LLSQALLPSLFQKLLWSQLSGGLIPTR-WLDFAAYSALRASRGRESDAPTQKSL 118
DB 21 LLSPMVGLSLR-----AWSRLR-GLGPLEPWLVEAVKGAAL----- 55
QY 119 SYTAGLFAKTRVWVSTLALRGTPVAVLVLRLEVKLAQERALDSAP-----T 168

```

Db 21 LLSVPMGLLSR-----TWSRLR-GLGLEPWLVEAVKGAAL----- 55
QY 119 SYTAAGLFAKTRVNSTLARGCTVAVLRLVLEKKAQERALDSAP-----T 168
Db 56 --VEAGLEGEAR-----TFPLA--IHTPMGRPEBEAEDSGGFGEDRETGLGKT 100
QY 169 FLLEQQLMGVLLPSSLAQGLVSHRELDSGGLSVQSLGNFKVVSYLLNPSYL----- 223
Db 101 SSSLPEAWG--LLDD--DDGMVGEREATSVPRGQSQFADGQ---RAPLSFSLIRTLQ 152
QY 224 --DYLQGLGLQSSAGSQGVGFRILT-----PESCYLESDGCHPQPLRAEMATA--- 273
Db 153 GSDKNP--GEKAELEGVAREEGVKNFSPYPSHRECCPAVEEEDDEAAVKAEHRTTSA 210
QY 274 -----WRCP-----PLSTEGLPETHHMRMLVFLQNOGQDLPTLQDNGYHSLE 320
Db 211 LSPGSKPTWVSCPEERENQATEDKRTERSKGARKTSVSPSSGSDPRSWYRGEASEE 270
QY 321 BEHNLRLMDPOHCTDNPAAQVSPAADRPEPTKPELVIOBVSQSPQSSSLFCBLPVEKE 380
Db 271 KE-----EKAHKTGKEAA-----PGQSSAP-----AQLQLKSWWCQPSDEE 311
QY 381 CEEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILGAPSLDEASSDSESDWGEPEDD 440
Db 312 GEVKALGAARKDGEAECPCIPPPSAFLKAWVYMPG-----EDTEEEDEE---DESDS 364
QY 441 GFDGSLSESDVEQDSGLHL--WNSFHSVDYKPNFTATIO-----TAARTAPRDP 493
Db 365 GSDEEGEAREASSTPATGVFLKSH-----VTPGEDTEEEDESDTGSADERAE 417
QY 494 DSGTSWSGSCGVGQCEGLPETPDH-----SSGEEDDWEPSSADEAENLKLWNSFCHSE 547
Db 418 TSASTPPASAFKAWVYRPGEDTEEEDEDVDSKEDDSEALGEAES----- 466
QY 548 DPNYLLNFKAPQPSG-----KNWKGQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNC 603
Db 467 DPH-----FSHPDQRAHFRGMYRPGKETE---BAED-- 498
QY 604 PGCGLGEA-----LAGER-----YTHIKRK 623
Db 499 -----WGEAEPGCFRVAIVYVPEKEPPPPWAPRLPLRLQRLKRPETPHDPPETPLKAR 554
QY 624 KVTFLVEVTEYIIS-----GDEDRKGPWEFARDGCFQKRIQTEVAIGYCLAPEHREK 678
Db 555 KVPFSEKVTVHFLAWAGPAQAARQGPWEQLARDRSFARRITQAEELSPLCPAARAR 614
QY 679 MFNRLR 684
Db 615 AWARLR 620

RESULT 14

Q60465 PRELIMINARY; PRT; 590 AA.
AC Q60465;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gadd34 protein.
GN Name=Gadd34;
OS Cricetus longicaudatus (long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10030;
RN [1]
RP MEDLINE=94187707; PubMed=8139541;
RA Zhan Q., Lord K.A., Alamo I.Jr., Hollander M.C., Carrier F., Ron D.,
RA Kohn K.W., Hoffman B., Liebermann D.A., Fornace A.J.Jr.;
RT "The gadd and Myd genes define a novel set of mammalian genes encoding
RT acidic proteins that synergistically suppress cell growth.";
RL Mol. Cell. Biol. 14:2361-2371(1994).

DR EMBL; L28147; AAA36983.1; -.
DR PIR; A56535; A56535.
SQ SEQUENCE 590 AA; 64527 MW; B3D879BDACBAE6D6 CRC64;
Query Match 4.6%; Score 171.5; DB 2; Length 590;
Best Local Similarity 18.1%; Pred. No. 0.03; Indels 187; Gaps 18;
Matches 100; Conservative 68; Mismatches 198;
QY 250 PESCYLSEDCGCHPQPLRAEMASATAWRRCPPL-----STEGLPETHHRR 292
Db 103 PETILGLSDDDKQGDGPREQGRAHTAGLPILLSPGLQSDKSLGEVVGEGVELAYPT 162
QY 293 MRMLVFLQNOGQDLPTLQDNG-----YHSLREEH 323
Db 163 SHW-----EGCSEEBEDGETVKAFRASADSPGHKSSTSVYCPGEAEHQATEKQ 213
QY 324 NLLRMDP-----QHCTDNPAAQVSPAADREP-----TEKKPELV 358
Db 214 TENKADPPSPSGSHSRAWYCSKQGEA-----DPEHRAKGYQLCQNAEAESEBEAK 267
QY 359 IOEVSQSPQSSSL--FCLEPVEKECEDHTNATDLSDRGESLPVSTRPVCSNKLIDYILG 416
Db 268 VSSLVSSGNAFLKAWVYRPGEDTDDDDSDWGAEEGKALSSPTSPE-----HDFLKA 322
QY 417 GAPSDLEASSDSESDWGEPEDDGFDSDGSLSESDVEQDSGLHLMNSFHSVDYKPNQ 476
Db 323 WYVRPGEDTDDDDSDWGSABE-----EGKALSSPTSPEHDFLKA-----VYRPG 369
QY 477 FTATIQTAARIAPRDSDSGTSWSGSCGVGSCQEGPLPET----- 516
Db 370 DT-----EDDQSDWGAEEKDGLAQTTFATPHTSAFLKTVWCPCGEBTDDDC 416
QY 517 ---PDHSSGEEDWEPSS-----ADEAENLKLWNSFCHSEDPYLLNFKAPQPSGK 564
Db 417 EVVVPDSEADPKSPSHEAQCLPCEQTEGL-----VEAHSFLFOVAYFLPCEKAP 470
QY 565 NWKGQDSKASSEVTVAFSGHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRK 624
Db 471 PW-----TAPKPLRLQRLTLTRPTQ---DQDPETP-----LRARK 505
QY 625 VTFLVEVTEYIIS-----GDEDRKGPWEFARDGCFQKRIQTEVAIGYCLAPEHREK 679
Db 506 VHFSENVTVHFLAWAGPAQAARQGPWEQLARDRSFARRIAQAEELKGLPYLTPAFARA 565
QY 680 FNRLRTEKDLLL 692
Db 566 WARLGNPSLPAL 578

RESULT 15

Q9VB94 PRELIMINARY; PRT; 676 AA.
AC Q9VB94;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE CG6296-PA (LP07116p).
GN ORENAMES=CG6296;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Baou A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houshun K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celinker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RN SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

RN [7]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dreamek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacieb J., Paregas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003758; AAF56648.1; -;
DR EMBL; AY118607; AM49976.1; -;
DR HSP; PS4318; I8U8.
DR FlyBase; FBgn0039470; CG6296.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0008970; F: phospholipase A1 activity; IEA.
DR GO; GO:0006629; P: lipid metabolism; IEA.
DR InterPro; IPR002334; Dol/Ves_allerg.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00151; Lipase; 1.
DR PRINTS; PR00825; DOLALLERGEN.
DR PRINTS; PR00821; TAGLIPASE.
SQ SEQUENCE 676 AA; 72427 MW; 6A707E13C7AC5CCC CRC64;

Query Match 4.0%; Score 148.5; DB 2; Length 676;
Best Local Similarity 21.0%; Pred. No. 0.93;
Matches 115; Conservative 62; Mismatches 195; Indels 175; Gaps 27;

QY 78 WSQISG-GLIPTWILFPAASYSALRASGRRESDATPVQKLSYTAAGLPAKTRVVSTLA 136
DB 128 WFQYGDYNNMIAVDWL-----RGR-----SLEY-----ASSVA 154

QY 137 LARG-GTPVAVLRLVLEVKLKAQERALDSAAPTFLLE---QQLWGVLLPSSLQAQIVSH 192
DB 155 GAFGAGKKVAALV-----DFLVEGYGMSLDLTLVGFSLGAVHAGH 195

QY 193 --RELSSSGPLSVOSLGNFKVSVLLNP-----SYLDYLPQ-----GLRCCSSAGGGQ 241
DB 196 TAKQVNSGKVG-----KVVG--LDPASPLISYNTKEKLSDDALYVESIQNGA 243

QY 242 FVGF-RTLTPESCYL-----SEDGC-----HPQPLRAEMSATARRCPPLSTGLPE 287
DB 244 ILGFGQPIGKASFYMGGRSQPCGIDITGSCSHTRAVLYVVEALLWNFPKICSSVD 303

QY 288 IHRR-----NRWLVLFP-----NQQLDPTLDQDNGYHS 318
DB 304 ANKNCNTYSSVFMGASINFFVAEGIFYVPVNKESPYGLGELNSGGEATGTPTSTT 363

QY 319 LE-EHNLLRMDPQHCTDNPQAVSPAADPEPEKPKPELVIOEVSGSPGSSLFCPLPV 377
DB 364 VDGEDESTEVSSTTTDKPEESTTEPEEDSTTNGKPE-ESSTTTTEQPEDSTTTTPEPI 422

QY 378 EK--ECEEDHTNATDLSDRGESLPVTRPVCSNKLIDYILGGARSDLEASDSESDWG- 434
DB 423 DSTTEAPEDESTTSSTPDGGEQSTTEPETTEKPEE--TSTSPIDTSTTTKEPEVDST 480

QY 435 -----EPEDD-----GFDSD-----GSLSESDVEQDSEGLHWNFSHSDPY 472
DB 481 TPKESESTTSPEDDDTTSAPCEDDDTTAAPGDDEETTEDPEETTTSPSADDDSTTE 540

QY 473 KPNFTATICTAARIAP-----RDPSSGTSWSGCGVSGCQEGPLPETPHSHSGEEDWE 528
DB 541 EPEEDTTT--TKKVPSPSTTTEPEEDSTT-----EVPSESTTEPDKE 581

QY 529 PSADAE 535
DB 582 TSTTNAE 588

Search completed: September 16, 2005, 10:46:02
Job time : 75.2452 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 08:08:59 ; Search time 1425.17 Seconds
(without alignments)
12220.183 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2942

Sequence: 1 attttggcttcgttcac.....aaattatgtgctctatctg 2942

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : N Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2942	100.0	2942	12	ADQ88457
2	2934	99.7	2942	4	AAH18699
3	2185.8	74.3	2324	3	AAC98158
4	1324	45.0	5468	12	ADQ88459
5	828.6	28.2	883	4	AAH92389
6	738	25.1	772	4	AAH06623
7	515	17.5	543	4	AAH11666
8	265.8	9.0	561	6	ABL38451
9	218	7.4	218	3	AAC29803
10	199	6.8	706	6	ABK44494
11	198.6	6.8	399	8	ABX52167
12	195	6.6	195	3	AAC06938
13	187.4	6.4	628	6	AAH61819
14	162	5.5	173	10	ADD49398
15	61.8	2.1	2000	8	ADA71938
16	46.4	1.6	8346	6	ABK28328
17	45.8	1.6	15881	12	ADQ19901
18	44.2	1.5	193672	10	ADL13570
19	43.6	1.5	778	6	ABQ15588
20	43.6	1.5	778	6	ABQ15589

21	43.4	1.5	5986	6	ABK28397	Abk28397 DNA trans
22	43.2	1.5	6794	6	ABK31265	Abk31265 Signal tr
23	43.2	1.5	6794	6	ABL70220	AbL70220 Chemically
24	43.2	1.5	6794	6	AAH61175	AAH61175 Human gen
25	43.2	1.4	110000	12	ADH69807_0	ADH69807 Human vbe
26	42.2	1.4	267156	6	ABL68560	AbL68560 Kidney ca
27	42	1.4	2000	8	ADA71938	Ada71938 Rice gene
28	42	1.4	15649	4	AAH45397	AAH45397 Chemically
29	42	1.4	15649	6	ABK28242	Abk28242 DNA trans
30	41.6	1.4	1863	4	AAH16572	AAH16572 Human CDN
31	41.6	1.4	5983	10	ADA13381	Ada13381 Human int
32	41.6	1.4	32640	13	ACN37228	ACN37228 Human per
33	40.8	1.4	19236	6	ABN80227	ABN80227 Human che
34	40.6	1.4	629	13	ADQ51068	AdQ51068 Novel can
35	40.6	1.4	8093	8	ABZ10177	Abz10177 Haematopo
36	40.6	1.4	8093	10	ADB54267	ADB54267 Pretreate
37	40.6	1.4	8093	10	ADH84189	ADH84189 Human lym
38	40.6	1.4	8093	13	ADH89565	ADH89565 Oligonuc
39	40.4	1.4	309	5	ABV61252	ABV61252 Human pro
40	40.4	1.4	6259	6	ABL32290	AbL32290 Human imm
41	40.2	1.4	461	5	ADL43330	AdL43330 Human ova
42	40.2	1.4	1293	13	ADS09901	ADs09901 Human the
43	40.2	1.4	1460	6	AAH62717	AAH62717 cDNA sequ
44	40.2	1.4	1560	6	ABX04170	ABx04170 Human mRN
45	40.2	1.4	1560	6	ABK84105	AbK84105 Human CDN
46	40.2	1.4	1560	6	ABN95913	ABn95913 Gene #241
47	40.2	1.4	1560	8	ABX55995	ABx55995 Gene enco
48	40.2	1.4	1560	10	ADC35146	ADC35146 Human bre
49	40.2	1.4	1560	11	ADL31851	ADl31851 Human CDN
50	40.2	1.4	1560	13	ACN39529	ACn39529 Tumour-as
51	40.2	1.4	1647	5	AAH80933	AAH80933 DNA enco
52	40.2	1.4	2410	5	ADL63044	ADl63044 Human ova
53	40.2	1.4	2908	12	ADQ29596	AdQ29596 Human col
54	40.2	1.4	3585	10	ADG32741	ADg32741 Human DNA
55	40.2	1.4	4252	12	ADH77169	ADe77169 Human CDN
56	40.2	1.4	6048	4	AAH46614	AAH46614 Tumour su
57	40.2	1.4	19345	6	ABN80017	ABn80017 Human che
58	40	1.4	752	6	ABQ54799	ABq54799 Human ova
59	40	1.4	6816	12	ADQ24856	ADq24856 Human sof
60	40	1.4	7341	6	AAH61394	AAH61394 Human gen
61	40	1.4	18624	6	ABL33703	AbL33703 Human imm
62	40	1.4	61020	4	AAH46787	AAH46787 Tumour su
63	39.8	1.4	6167	6	ABK31407	AbK31407 Signal tr
64	39.8	1.4	6167	6	ABL70368	AbL70368 Chemically
65	39.8	1.4	6167	6	AAH61330	AAH61330 Human gen
66	39.8	1.4	6169	4	AAH46369	AAH46369 Tumour su
67	39.8	1.4	6169	6	ABN80096	ABn80096 Human che
68	39.8	1.4	16766	6	ABL34157	AbL34157 Human imm
69	39.6	1.3	20486	7	ADS99871	ADs99871 Bistulphit
70	39.6	1.3	20486	7	ADH56232	ADH56232 AmEPV apo
71	39.4	1.3	453	6	ABL56232	ABl56232 AmEPV apo
72	39.4	1.3	3523	8	ABZ10170	Abz10170 Haematopo
73	39.4	1.3	6523	10	ADH84216	ADH84216 Human lym
74	39.4	1.3	6523	13	ADH89620	ADH89620 Oligonuc
75	39.4	1.3	50000	6	ABL56202	ABl56202 AmEPV gen
76	39.2	1.3	10047	8	ACF73323	ACf73323 Staphyloc
77	39.2	1.3	5311	6	ABL33019	AbL33019 Human imm
78	39.2	1.3	11147	8	ABZ10153	Abz10153 Haematopo
79	39.2	1.3	14147	10	ADH54325	ADH54325 Pretreate
80	39.2	1.3	14147	10	ADH84163	ADH84163 Human lym
81	39.2	1.3	14147	13	ADH89523	ADH89523 Oligonuc
82	39.2	1.3	23439	2	AAV74349	AAV74349 Staphyloc
83	39.2	1.3	56153	4	AAH46794	AAH46794 Tumour su
84	39.2	1.3	61648	12	ADQ97663	ADQ97663 Human can
85	39	1.3	387	5	ADL71793	ADl71793 Human ova
86	39	1.3	387	5	ADL36946	ADl36946 Human ova
87	39	1.3	2990	6	ABQ54800	ABq54800 Human ova
88	39	1.3	8093	4	AAH46435	AAH46435 Tumour su
89	39	1.3	8093	6	ABK33973	ABk33973 Human DNA
90	39	1.3	8093	6	ABL92236	ABl92236 Chemically
91	39	1.3	8093	6	ABL49331	ABl49331 Human MLH
92	39	1.3	8093	8	ABZ10031	ABz10031 Haematopo
93	39	1.3	8093	8	ADA20360	ADA20360 Prostate

94 39 1.3 8093 8 ADA84167
 95 39 1.3 8093 10 ADB54139
 96 39 1.3 8093 10 ADE84113
 97 39 1.3 8093 13 ADS89291
 98 39 1.3 19087 6 ABL32792
 c 99 38.8 1.3 757 2 AAX55996
 100 38.8 1.3 6171 6 ABL33010

ADA84167 Human ren
 Adb54139 Pretreate
 Ade84113 Human lym
 Ad889291 OligonucL
 Ab132792 Human imm
 Aax55996 Human cDN
 Ab133010 Human imm

ALIGNMENTS

RESULT 1

ADQ88457
 ID ADQ88457 standard; cDNA; 2942 BP.

XX

AC ADQ88457;

XX 07-OCT-2004 (first entry)

XX Human GADD34-like (GADD34L) cDNA.

XX Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;
 KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;
 KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human; gene;
 XX ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 PH 407..2548
 FT CDS
 FT /*tag= a
 FT /product= "Human GADD34-like (GADD34L) protein"

XX US2004142345-A1.

PN 22-JUL-2004.

PD 28-AUG-2003; 2003US-00650482.

XX 06-SEP-2002; 2002US-0408679P.

PR (ROND/) RON D.

PA (JOUS/) JOUSSE C.

XX Ron D, Jousse C;

XX WPI; 2004-552556/53.

DR P-PSDB; ADQ88458.

DR GENBANK; AK027650.

XX Screening test substances for preventing or treating disease involving
 PT oxidative stress, by testing test substances for its ability to inhibit
 PT activity of GADD34L and identifying test substance that inhibits activity
 PT of GADD34L.

XX Disclosure; SEQ ID NO 1; 30pp; English.

XX The present invention relates to a method of screening several test
 CC substances for preventing or treating diseases involving oxidative stress
 CC such as neuronal ischaemia, heart ischaemia, renal damage induced by
 CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.
 CC The method involves testing the test substances for its ability to
 CC inhibit the activity of GADD34-like (GADD34L), also referred to as
 CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the
 CC test substance which inhibits the activity of GADD34L. The present
 CC sequence is human GADD34L cDNA.

SQ Sequence 2942 BP; 732 A; 739 C; 725 G; 746 T; 0 U; 0 Other;

Query Match 100.0%; Score 2942; DB 12; Length 2942;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCCCTTCGGGTATCGCCTTG	60
Db	1	ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCCCTTCGGGTATCGCCTTG	60
Qy	61	CTCAGGGGCTTTTCAACCCCTCTGTAGTCGGAACACCATCGCCGAGGCGGTGGGGGAGCT	120
Db	61	CTCAGGGGCTTTTCAACCCCTCTGTAGTCGGAACACCATCGCCGAGGCGGTGGGGGAGCT	120
Qy	121	CCTATCCATGGTGTGAAGCGTCGAGCGGAGCTAGGGAACCTTCTCCCGCCAGATGGA	180
Db	121	CCTATCCATGGTGTGAAGCGTCGAGCGGAGCTAGGGAACCTTCTCCCGCCAGATGGA	180
Qy	181	AGTCGCATCAGTCGCGCGCTATTTCGCGGGGCTGTCTTCCCTGTGTCTGCGCCCGCTG	240
Db	181	AGTCGCATCAGTCGCGCGCTATTTCGCGGGGCTGTCTTCCCTGTGTCTGCGCCCGCTG	240
Qy	241	CCGCATTTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTCGAGCAGCGAC	300
Db	241	CCGCATTTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTCGAGCAGCGAC	300
Qy	301	GCACCGCTGGGCAAGCCGAGACTCTGTAGGCTTCTTCGAATCCCGTCGACCTCAGC	360
Db	301	GCACCGCTGGGCAAGCCGAGACTCTGTAGGCTTCTTCGAATCCCGTCGACCTCAGC	360
Qy	361	CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAGAGAGATGAGCCGGGAC	420
Db	361	CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAGAGAGATGAGCCGGGAC	420
Qy	421	AGCGGATTCGCGGAACGGCTTGGCCCTCGGGGGGCTTCGGTTCCTGGCCACCCCTTTT	480
Db	421	AGCGGATTCGCGGAACGGCTTGGCCCTCGGGGGGCTTCGGTTCCTGGCCACCCCTTTT	480
Qy	481	CCCTCGGGATTCGCAAGCAGGCTTCTTAAGTTCCCGACGCTTCGCGCGGAAACTC	540
Db	481	CCCTCGGGATTCGCAAGCAGGCTTCTTAAGTTCCCGACGCTTCGCGCGGAAACTC	540
Qy	541	CGGGAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAGTTACTTGAGCGAA	600
Db	541	CGGGAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAGTTACTTGAGCGAA	600
Qy	601	ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATGCTTCAAGAGTGTCTTAATTTGAG	660
Db	601	ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATGCTTCAAGAGTGTCTTAATTTGAG	660
Qy	661	CCAACTTTTCGGTGGAAATGTTTCCGACCCAGATGGCTAGATTTTCTGGAGTCTACAGCG	720
Db	661	CCAACTTTTCGGTGGAAATGTTTCCGACCCAGATGGCTAGATTTTCTGGAGTCTACAGCG	720
Qy	721	CCTGAGAGCCTTGAAGGAGCGGAGAAACAGCGCCGCCACAGCGCAGAGAAATCTTTGAG	780
Db	721	CCTGAGAGCCTTGAAGGAGCGGAGAAACAGCGCCGCCACAGCGCAGAGAAATCTTTGAG	780
Qy	781	TTGCTCTCAGCTCAGTCTTCCAGACCCCTCGGTCCAGTCCCTTGAATTTGGCTAGAGGA	840
Db	781	TTGCTCTCAGCTCAGTCTTCCAGACCCCTCGGTCCAGTCCCTTGAATTTGGCTAGAGGA	840
Qy	841	GGGATCCATGGCAATACCTCGCCCGCAGACCTTAAATTTGGAGCTTAAGCCAGGGAAG	900
Db	841	GGGATCCATGGCAATACCTCGCCCGCAGACCTTAAATTTGGAGCTTAAGCCAGGGAAG	900
Qy	901	TGCTTTTGGACCTTCGACGACAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960
Db	901	TGCTTTTGGACCTTCGACGACAGAGCTTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT	960
Qy	961	GTGCTCCAGTAGCTTCAATTCGGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC	1020
Db	961	GTGCTCCAGTAGCTTCAATTCGGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC	1020
Qy	1021	TGGGCTCTTAACATTCACCGCATAGACAATTTTCAGTGTGTATCTCTATTTTGTCTGAACCC	1080
Db	1021	TGGGCTCTTAACATTCACCGCATAGACAATTTTCAGTGTGTATCTCTATTTTGTCTGAACCC	1080

1081 TTCTACTGAGCTGCTTTCTTAGCTAGAAAGTCAGCTATCAGAACAGTGTGAAATAG 1140
 1081 TTCTACTGAGCTGCTTTCTTAGCTAGAAAGTCAGCTATCAGAACAGTGTGAAATAG 1140
 1141 CGAGGTAGTCGGCTTCAGACACTAAACCCAGAGAGAGCTGCTGAGAGAGACCATTTG 1200
 1141 CGAGGTAGTCGGCTTCAGACACTAAACCCAGAGAGAGCTGCTGAGAGAGACCATTTG 1200
 1201 TCATCCCCAGCGCTGAGTGCAGAACTCATTTCCGGCCTCGTGGCAGGAGATGTCACCTCT 1260
 1201 TCATCCCCAGCGCTGAGTGCAGAACTCATTTCCGGCCTCGTGGCAGGAGATGTCACCTCT 1260
 1261 TTCTACGGAAGCCCTACAGAAATTCACCAATTCGATGCAAAACCGCTGGAATTCCTTCA 1320
 1261 TTCTACGGAAGCCCTACAGAAATTCACCAATTCGATGCAAAACCGCTGGAATTCCTTCA 1320
 1321 ACAGGCTAACAGGGGCAAGATTTACCAACCCCTGACCCAGGATATGGCTACCCAGCCT 1380
 1321 ACAGGCTAACAGGGGCAAGATTTACCAACCCCTGACCCAGGATATGGCTACCCAGCCT 1380
 1381 GGAGGAGGAACACAGCCTTCTCCGGATGGATCCAAAACACTGCAGAGATACCCCAACACA 1440
 1381 GGAGGAGGAACACAGCCTTCTCCGGATGGATCCAAAACACTGCAGAGATACCCCAACACA 1440
 1441 GTTTGTTCTGCTGCTGGAGACATTCCTGGAAACACCCAGGAATCCACTGAAGAAAAT 1500
 1441 GTTTGTTCTGCTGCTGGAGACATTCCTGGAAACACCCAGGAATCCACTGAAGAAAAT 1500
 1501 AGAATTATTACTACAGAGTTCCACTGCTTTGGAGAGAGAGCCCTCTGAGGGCTG 1560
 1501 AGAATTATTACTACAGAGTTCCACTGCTTTGGAGAGAGAGCCCTCTGAGGGCTG 1560
 1561 TCCATCTAGTCAGATACCTATGGAAAGAGAGCTGGAGAGGCGGAAATAGTGTGTA 1620
 1561 TCCATCTAGTCAGATACCTATGGAAAGAGAGCTGGAGAGGCGGAAATAGTGTGTA 1620
 1621 TTACTCATACCTAGAGGTGACCTTCCATTTCTGCCAGACAGCTTGTAGTAAACAACT 1680
 1621 TTACTCATACCTAGAGGTGACCTTCCATTTCTGCCAGACAGCTTGTAGTAAACAACT 1680
 1681 GATAGATTATTTTGGAGGTGATCCAGTGACCTGGAACCAAGTTCGTGATCAGAGG 1740
 1681 GATAGATTATTTTGGAGGTGATCCAGTGACCTGGAACCAAGTTCGTGATCAGAGG 1740
 1741 TGAGGATTGGATGAGGAAGCTGAGGATGATGTTTTCATAGTATAGTCACTGTGAGA 1800
 1741 TGAGGATTGGATGAGGAAGCTGAGGATGATGTTTTCATAGTATAGTCACTGTGAGA 1800
 1801 CTCAGACCTTGAACAGACCCCTGAGAGGCTTCACTTTGGAACTCTTTCTGAGGTAGA 1860
 1801 CTCAGACCTTGAACAGACCCCTGAGAGGCTTCACTTTGGAACTCTTTCTGAGGTAGA 1860
 1861 TCCTTATATCCCGAATTTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTCTGTA 1920
 1861 TCCTTATATCCCGAATTTTACAGCAACAAATTCAGACTGCTGCCAGAAATGTTCTGTA 1920
 1921 AGAGCCTTCTGATTCAGAGAAGGATTTCTGCGCAAGTCTGATCTAGAGAATTCCTCCA 1980
 1921 AGAGCCTTCTGATTCAGAGAAGGATTTCTGCGCAAGTCTGATCTAGAGAATTCCTCCA 1980
 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCAGTATGTTCTGGGAGGAAGATGACTGGAAATC 2040
 1981 GTCTGGAAGCCTTCTGAGACCCCTGAGCAGTATGTTCTGGGAGGAAGATGACTGGAAATC 2040
 2041 TAGTGCAGATGAGCAGAGAGTCTCAAACTGTGNACTCATTTCTGTAATTCGTGATGCC 2100
 2041 TAGTGCAGATGAGCAGAGAGTCTCAAACTGTGNACTCATTTCTGTAATTCGTGATGCC 2100
 2101 CTACAAACCTTTAAATTTTAAAGCTCTTTTCAAAACATCAGGGGAAAATGAGAAAGGCTG 2160
 2101 CTACAAACCTTTAAATTTTAAAGCTCTTTTCAAAACATCAGGGGAAAATGAGAAAGGCTG 2160
 2161 TCGTGAATCAAGAGCCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT 2220

2161 TCSTGACTCAAGACCCCATCTGAGTCCATTTGTGGCCATTTCTGAGTGTCAACCTTACT 2220
 2221 TTCTTTGAAGTGCAGCTGTTGGGAGCCAAAGAAATGGAATGTCCAGACTCGGTACAGCG 2280
 2221 TTCTTTGAAGTGCAGCTGTTGGGAGCCAAAGAAATGGAATGTCCAGACTCGGTACAGCG 2280
 2281 TGAGCTTTCTTTCTGGAGGAAGACACACATGTCAAAAGAAAAGGTAAACCTTTCTTTGA 2340
 2281 TGAGCTTTCTTTCTGGAGGAAGACACACATGTCAAAAGAAAAGGTAAACCTTTCTTTGA 2340
 2341 AGAAGTTACTGAGTATTTATATAGTGTGATGAGGATCGCAAAGGACCATGGGAGAAAT 2400
 2341 AGAAGTTACTGAGTATTTATATAGTGTGATGAGGATCGCAAAGGACCATGGGAGAAAT 2400
 2401 TGCAAGGATGGATGCAAGTTTCAGAAACGAAATTCAGAAACAGAAATGCTATTGATA 2460
 2401 TGCAAGGATGGATGCAAGTTTCAGAAACGAAATTCAGAAACAGAAATGCTATTGATA 2460
 2461 TTGCTTGACATTTGAAACACAGAGAAAGAAATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520
 2461 TTGCTTGACATTTGAAACACAGAGAAAGAAATGTTTAAATAGACTCCAGGGAACATGCTTCAA 2520
 2521 AGAATTTAATGTTCTCAGCAATGTTGAGTTGGCAGCTGTAGTCTTAGCTAGCATACAC 2580
 2521 AGAATTTAATGTTCTCAGCAATGTTGAGTTGGCAGCTGTAGTCTTAGCTAGCATACAC 2580
 2581 TACCTCTTACCTGAGAGGTGCTTTTAAAAACAAATCTTCGAGCTGTCTTTTGACATTT 2640
 2581 TACCTCTTACCTGAGAGGTGCTTTTAAAAACAAATCTTCGAGCTGTCTTTTGACATTT 2640
 2641 TTTTCTTTAGAGGAATGTAACCTGAGTCTAGTTTAAATTTTGTGCAACATATCCC 2700
 2641 TTTTCTTTAGAGGAATGTAACCTGAGTCTAGTTTAAATTTTGTGCAACATATCCC 2700
 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATT 2760
 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATT 2760
 2761 CTAACTCTCCCTTTTGTGATTTAGTTGATGCTTTTAAATGCTTTCCTGCTCATGAG 2820
 2761 CTAACTCTCCCTTTTGTGATTTAGTTGATGCTTTTAAATGCTTTCCTGCTCATGAG 2820
 2821 GTGAAAAGGGACCTTTTTCAGTTGTCATTTTGCATTTTCAAACTTTATTTCTTGGAAA 2880
 2821 GTGAAAAGGGACCTTTTTCAGTTGTCATTTTGCATTTTCAAACTTTATTTCTTGGAAA 2880
 2881 ACAATATTTATAGGCTTTAAAGCCCATTTTCATTTCTAAATTAATGTTGCTCATC 2940
 2881 ACAATATTTATAGGCTTTAAAGCCCATTTTCATTTCTAAATTAATGTTGCTCATC 2940
 2941 TG 2942
 2941 TG 2942

RESULT 2

AAH18699

ID AAH18699 standard; cDNA; 2942 BP.

XX AAH18699;

AC AAH18699;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:18964.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX OS

XX EP1074617-A2.

XX FN

XX 07-FEB-2001.

XX PD

XX 28-JUL-2000; 2000EP-00116126.
 XX 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 XX Claim 8; SEQ ID NO 18964; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 XX Sequence 2942 BP; 731 A; 741 C; 726 G; 744 T; 0 U; 0 Other;
 SQ

Db GCCACCGCTGGCAAGCCGAGACTCTGTAGCTTCTCCGAATCCCGTGCACCTCCAGC 360
 QY CGCTGAGCGCGCGCCCTTACCTGAGAGACTGTCAAGAAAAAGAGATGAGAGCGGGAC 420
 Db CGCTGAGCGCGCGCCCTTACCTGAGAGACTGTCAAGAAAAAGAGATGAGAGCGGGAC 420
 QY AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGGCTTCGGGTTCTGGCCACCTTTT 480
 Db AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGGCTTCGGGTTCTGGCCACCTTTT 480
 QY CCCTCGCGGATCGCAAGCAGGCTTCTTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTC 540
 Db CCCTCGCGGATCGCAAGCAGGCTTCTTAAGTTTCCCGACGCTCTTGGCCCGGAAAACTC 540
 QY CGGAAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAAGTTTACTTGAGCGAA 600
 Db CGGAAACCCACACTGCTTCTGCGCCAGCCGAGACTCGGGTCAAGTTTACTTGAGCGAA 600
 QY ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGTTTCTGAGAGTGTCTAATTTGAG 660
 Db ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGTTTCTGAGAGTGTCTAATTTGAG 660
 QY CCAACTTTTTCGGTGAATGTTTCCGACAGATGCTAGATTTTCTGAGGCTCTACAGCGC 720
 Db CCAACTTTTTCGGTGAATGTTTCCGACAGATGCTAGATTTTCTGAGGCTCTACAGCGC 720
 QY CTTGAGAGCTTGAAGGAGCGGAGAAACAGCGCCCGCCACAGCGAGAGAACTTTTGG 780
 Db CTTGAGAGCTTGAAGGAGCGGAGAAACAGCGCCCGCCACAGCGAGAGAACTTTTGG 780
 QY TTGCTCAGCTCAGCTCCTCAGACCTCGCTCGCTCAGCTCCCTTCAATTTGGCTTAGAGGA 840
 Db TTGCTCAGCTCAGCTCCTCAGACCTCGCTCGCTCAGCTCCCTTCAATTTGGCTTAGAGGA 840
 QY GGGGATCCACTGGCAATPACTCGCCCGCCAGACCTTAAATTTGAGGCTTAAAGCCAAAGGAAG 900
 Db GGGGATCCACTGGCAATPACTCGCCCGCCAGACCTTAAATTTGAGGCTTAAAGCCAAAGGAAG 900
 QY TGCTTTGAGCGCTGACAGCAGAGCTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960
 Db TGCTTTGAGCGCTGACAGCAGAGCTTTCTTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960
 QY GTTCCCGAGTAGCTTCAATCCCGTCTGTACTCTAAACCGGAACTTGGCTTTCGCCCTC 1020
 Db GTTCCCGAGTAGCTTCAATCCCGTCTGTACTCTAAACCGGAACTTGGCTTTCGCCCTC 1020
 QY TGGGCTCTTAAACATTCACGCAATAGCAATTTCAAGTGTGGTATCTTATTTGCTGAACCC 1080
 Db TGGGCTCTTAAACATTCACGCAATAGCAATTTCAAGTGTGGTATCTTATTTGCTGAACCC 1080
 QY TTCTTACTGAGCTGCTTTCTTAGGCTAGAGTCAAGTATCAGACAGTATGAGAAATAG 1140
 Db TTCTTACTGAGCTGCTTTCTTAGGCTAGAGTCAAGTATCAGACAGTATGAGAAATAG 1140
 QY CAGGTAGTGGCTTCCAGACACTAAACCCAGAGCAGCTGCTCTGAGAGAGGACCATTTG 1200
 Db CAGGTAGTGGCTTCCAGACACTAAACCCAGAGCAGCTGCTCTGAGAGAGGACCATTTG 1200
 QY TCATTCGCCAGCGCTGAGTCAAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACTCT 1260
 Db TCATTCGCCAGCGCTGAGTCAAGAACTCATTTCCGGCTCTGTGGCAGGATGTCCACTCT 1260
 QY TTCTTACGGAGGCTTACGAAATTTACCATCTTCGATCAAAACGGCTGGAATTTCTTCA 1320
 Db TTCTTACGGAGGCTTACGAAATTTACCATCTTCGATCAAAACGGCTGGAATTTCTTCA 1320
 QY ACAGGCTTAAACAGGGGCAAGATTTTACCCAGCTTGAACAGGCTGGAATTTCTTCA 1380
 Db ACAGGCTTAAACAGGGGCAAGATTTTACCCAGCTTGAACAGGCTGGAATTTCTTCA 1380
 QY GAGGAGGAAACAGAGCTTCTCGGATGGATTCAAAACACTGAGAGATTAACCAACACA 1440
 Db GAGGAGGAAACAGAGCTTCTCGGATGGATTCAAAACACTGAGAGATTAACCAACACA 1440

Db	1381	GGAGGAGGAA	CACAGCC	TTCTCTCGGATGGAT	TCCAAAACAT	CTGCAGAGANTAA	CCCAACACA	1444																				
Qy	1441	GTTTGT	TCTGCTCGAGACAT	TCTCGGAAACAC	CCAGGAAATCC	ACTGAAAGAAAAAT	1500																					
Db	1441	GTTTGT	TCTGCTCGAGACAT	TCTCGGAAACAC	CCAGGAAATCC	ACTGAAAGAAAAAT	1500																					
Qy	1501	AGAA	TTATTA	ACTACAGAGTT	CACTTGCTTT	TGGAAAGAGAGCC	TCTTGAGGGCTG	1560																				
Db	1501	AGAA	TTATTA	CTACAGAGTT	CACTTGCTTT	TGGAAAGAGAGCC	TCTTGAGGGCTG	1560																				
Qy	1561	TCCAT	CTAGTGAGAT	CCTATGGA	AAGGAGCC	TGGAGAGGGCC	GAATAGTGTAGTTGA	1620																				
Db	1561	TCCAT	CTAGTGAGAT	CCTATGGA	AAGGAGCC	TGGAGAGGGCC	GAATAGTGTAGTTGA	1620																				
Qy	1621	TTACT	CATAC	TAGAGGT	CACTTCC	ATTTCTG	CCAGACCA	GCTTGTAGTAA	1680																			
Db	1621	TTACT	CATAC	TAGAGGT	CACTTCC	ATTTCTG	CCAGACCA	GCTTGTAGTAA	1680																			
Qy	1681	GATAG	ATTATA	TTTTTGG	AGGTG	CATCCAG	TGAAC	CTGGA	AACAA	GTTCTG	ATTC	CAGAA	GG	1740														
Db	1681	GATAG	ATTATA	TTTTTGG	AGGTG	CATCCAG	TGAAC	CTGGA	AACAA	GTTCTG	ATTC	CAGAA	GG	1740														
Qy	1741	TGAG	GATTGG	GATGAG	GATGG	TTTTG	ATAGT	GATAG	CTCA	CTGT	CAGA	1800																
Db	1741	TGAG	GATTGG	GATGAG	GATGG	TTTTG	ATAGT	GATAG	CTCA	CTGT	CAGA	1800																
Qy	1801	CTCAG	ACCT	TGAA	CAGAC	CCTGA	AGGCT	TTCAC	CTTTG	GAAC	TCTTCT	G	CAGT	GTAGA	1860													
Db	1801	CTCAG	ACCT	TGAA	CAGAC	CCTGA	AGGCT	TTCAC	CTTTG	GAAC	TCTTCT	G	CAGT	GTAGA	1860													
Qy	1861	TCCTTA	TAAT	TCCC	CAGAA	CTTTAC	AGCA	CAATTC	CAG	CTGCTG	CCAGAA	TTGTTCT	CTG	A	1920													
Db	1861	TCCTTA	TAAT	TCCC	CAGAA	CTTTAC	AGCA	CAATTC	CAG	CTGCTG	CCAGAA	TTGTTCT	CTG	A	1920													
Qy	1921	AGAC	CTTCTG	ATTC	CAGAA	AGGAT	TTGTTCT	GCA	GTC	GCA	GTC	GAT	CTAG	AA	TTCC	TCC	TCCCA	1980										
Db	1921	AGAC	CTTCTG	ATTC	CAGAA	AGGAT	TTGTTCT	GCA	GTC	GCA	GTC	GAT	CTAG	AA	TTCC	TCC	TCCCA	1980										
Qy	1981	GTCTG	GAAG	CCTT	CTCAG	ACCC	CTGAG	CATAG	TCTT	CTGG	GAGAA	GATG	ACTG	AT	TGG	GAATC	2040											
Db	1981	GTCTG	GAAG	CCTT	CTCAG	ACCC	CTGAG	CATAG	TCTT	CTGG	GAGAA	GATG	ACTG	AT	TGG	GAATC	2040											
Qy	2041	TAGT	GCAG	ATGA	GCAG	AGTCT	CAAA	CTGTG	GA	ACTCA	TTCTG	TAAT	TCTG	AT	TCTG	AT	GACCC	2100										
Db	2041	TAGT	GCAG	ATGA	GCAG	AGTCT	CAAA	CTGTG	GA	ACTCA	TTCTG	TAAT	TCTG	AT	TCTG	AT	GACCC	2100										
Qy	2101	CTACA	CCTTTAA	ATTTAA	GGCTC	CTTTT	CAAA	CAT	CAGGG	AAAT	GAGAA	GGCTG	2160															
Db	2101	CTACA	CCTTTAA	ATTTAA	GGCTC	CTTTT	CAAA	CAT	CAGGG	AAAT	GAGAA	GGCTG	2160															
Qy	2161	TCGTG	ACT	CAAA	GACCC	CATCT	GAGT	CCAT	TGTG	GCC	ATTTCT	GAGT	GT	CACA	CC	TTTACT	2220											
Db	2161	TCGTG	ACT	CAAA	GACCC	CATCT	GAGT	CCAT	TGTG	GCC	ATTTCT	GAGT	GT	CACA	CC	TTTACT	2220											
Qy	2221	TTCTT	GTAAG	TG	CAG	CTGTT	TGGGG	GCC	AAGAA	G	TAAT	TG	TCC	GACT	CGGT	TAC	GCG	2280										
Db	2221	TTCTT	GTAAG	TG	CAG	CTGTT	TGGGG	GCC	AAGAA	G	TAAT	TG	TCC	GACT	CGGT	TAC	GCG	2280										
Qy	2281	TG	CGT	TTCT	TCTG	GAGG	AAGACA	CACA	CAT	GT	C	AA	AAG	AA	AA	GG	TAA	AC	TT	CT	TTGA	2340						
Db	2281	TG	CGT	TTCT	TCTG	GAGG	AAGACA	CACA	CAT	GT	C	AA	AAG	AA	AA	GG	TAA	AC	TT	CT	TTGA	2340						
Qy	2341	AGA	GT	TACT	CG	AT	TATAT	TAT	TAT	TAGT	GT	GAT	GAG	AT	CG	CA	AA	GGA	CC	AT	GG	GAG	AA	TT	2400			
Db	2341	AGA	GT	TACT	CG	AT	TATAT	TAT	TAT	TAGT	GT	GAT	GAG	AT	CG	CA	AA	GGA	CC	AT	GG	GAG	AA	TT	2400			
Qy	2401	TG	CA	AGG	GAT	TG	GAT	CT	CA	GA	AA	C	GA	AT	T	CA	GA	AA	C	GA	AT	T	CA	GA	AA	TT	2460	
Db	2401	TG	CA	AGG	GAT	TG	GAT	CT	CA	GA	AA	C	GA	AT	T	CA	GA	AA	C	GA	AT	T	CA	GA	AA	TT	2460	
Qy	2461	TTG	CT	TG	CA	AT	TTG	AA	CACA	GAG	AA	GA	AT	TG	TTTAA	T	TAG	AT	CT	CA	GG	GA	CA	TG	CT	TT	CA	2520
Db	2461	TTG	CT	TG	CA	AT	TTG	AA	CACA	GAG	AA	GA	AT	TG	TTTAA	T	TAG	AT	CT	CA	GG	GA	CA	TG	CT	TT	CA	2520

QY	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGGCTGTAGTCTCTAGCTAGCATACAC	2580
Db	2521	AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGGCTGTAGTCTCTAGCTAGCATACAC	2580
QY	2581	TACCTCTTTTACTGTAGAGGTGTCTTTTAAAAACAAATCTTGGCAGCTGTCTTTTGACATTT	2640
Db	2581	TACCTCTTTTACTGTAGAGGTGTCTTTTAAAAACAAATCTTGGCAGCTGTCTTTTGACATTT	2640
QY	2641	TTTTTTTTTAGAGGAAATGTAACCTTGGATCTAGTTTAAATTTTTTTTTTGGCAACATATCCC	2700
Db	2641	TTTTTTTTTAGAGGAAATGTAACCTTGGATCTAGTTTAAATTTTTTTTTTGGCAACATATCCC	2700
QY	2701	ACTCAGAAAAATTCAGGTTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT	2760
Db	2701	ACTCAGAAAAATTCAGGTTTTGAAGCCAGCCCTGATTAATGAAGATGAACCTAGTGTGATTT	2760
QY	2761	CTAATCCTCCCTTTTTTGGATTTAGTTGGATGTGCTTTTAAATGTCCTTTTAAATGTCCTTTGCAATGAG	2820
Db	2761	CTAATCCTCCCTTTTTTGGATTTAGTTGGATGTGCTTTTAAATGTCCTTTTAAATGTCCTTTGCAATGAG	2820
QY	2821	GTGGAAGGGGACCTTTTTCAGTTGTGCATTTTGCACCTTTCAAACTTATTTTCTTGAAA	2880
Db	2821	GTGGAAGGGGACCTTTTTCAGTTGTGCATTTTGCACCTTTCAAACTTATTTTCTTGAAA	2880
QY	2881	ACATATTTATAGGCTTAAAGCCATTTTTCATTTCTAAATCTAAATATGTGTGCCTATC	2940
Db	2881	ACATATTTATAGGCTTAAAGCCATTTTTCATTTCTAAATCTAAATATGTGTGCCTATC	2940
QY	2941	TG 2942	
Db	2941	TG 2942	
RESULT 3			
AAC98158			
ID	AAC98158 standard; cdna; 2324 BP.		
AC	AAC98158;		
XX			
DT	09-MAR-2001 (first entry)		
XX	Human colon cancer antigen nucleotide sequence SEQ ID NO:168.		
DE			
XX	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	identification; cytostatic; cardioactive; neuroprotective; vulnery;		
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;		
KW	nephrotropic; antiinfective; antibacterial; gene therapy; wound;		
KW	neural disorder; immune system disorder; muscular disorder;		
KW	reproductive disorder; gastrointestinal disorder; renal disorder;		
KW	infectious disease; cardiovascular disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200055351-A1.		
XX			
PD	21-SEP-2000.		
XX			
PF	08-MAR-2000; 2000WO-US005883.		
XX			
XX	12-MAR-1999; 99US-0124270P.		
PR			
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Rosen CA, Ruben SM;		
XX			
PI	WPI; 2000-587534/55.		
XX			
DR	P-PSDB; AAB53401.		
XX			
PT	Colon cancer associated gene sequences, referred to as colon cancer		
PT	antigens, useful for the treatment, prevention, and diagnosis of colon		
PT	disorders such as colon cancer.		
XX			

Claim 1; Page 595; 2104pp; English.

PS AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AB53234 to AB54006. The
CC human colon cancer antigens can have cytoskeletal, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antinefctive and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins may
CC also be used to prevent diseases such as neural disorders, immune system
CC disorders, muscular disorders, reproductive disorders, gastrointestinal
CC disorders, wounds, renal disorders, infectious diseases, and
CC cardiovascular disorders. AAC98764 to AAC98772 and AB54007 represent
CC sequences used in the exemplification of the present invention
XX
SQ Sequence 2324 BP; 615 A; 566 C; 562 G; 574 T; 0 U; 7 Other;

Query Match 74.3%; Score 2185.8; DB 3; Length 2324;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 2249; Conservative 6; Mismatches 36; Indels 6; Gaps 4;

QY 456 GCTTCGGGTTCTGGCCACCCCTTTTCCCTCGCGCATCGCAAGCGCTCTTCTAAGTTCC 515
DB 32 GCTCAGGGGCTTTTCAACCCCTGTGTCAGTCGCGCATCGCAAGCGCTCTTCTAAGTTCC 91
QY 516 CGACGCCCTCTGGCCCGGAAACTCCGGGAACCCACACTGCTGTTTCTCTGCCAGCCCG 575
DB 92 CGACGCCCTCTGGCCCGGAAACTCCGGGAACCCACACTGCTGTTTCTCTGCCAGCCCG 151
QY 576 AGACTCGGGTCAGTTACTGCGAAGCTGCTCTCCAGCTCTTGGCGCGCTCCCGGAT 635
DB 152 AGACTCGGGTCAGTTACTGCGAAGCTGCTCTCCAGCTCTTGGCGCGCTCCCGGAT 211
QY 636 TGCTTCAGAAAGTGCTAAATTTGGAGCCCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGC 695
DB 212 TGCTTCAGAAAGTGCTAAATTTGGAGCCCAACTTTTTCGGTGGAAATGTTTCCGACCAAGATGGC 271
QY 696 TAGATTTTGTGGAGCTACAGCGCCCTGAGAGCCCTGAGGGACGGGAGAAACAGCCG 755
DB 272 TAGATTTTGTGGAGCTACAGCGCCCTGAGAGCCCTGAGGGACGGGAGAAACAGCCG 331
QY 756 CCCACACAGCGCAGAAATCTTTAGTTCGCTGCAGCTCGACTCTCAGACCCCTCGGTCA 815
DB 332 CCCACACAGCGCAGAAATCTTTAGTTCGCTGCAGCTCGACTCTCAGACCCCTCGGTCA 391
QY 816 CCAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATFACTCGCCCCCAGACCTAA 875
DB 392 CCAGTCCCTTGAATTTGGCTAGAGGAGGGATCCACTGGCAATFACTCGCCCCCAGACCTAA 451
QY 876 AATTGGAGCTTAAGCCCAAGGAAGTGTCTTGGACCCCTGAGCAAGAGCTTTCTCTTAG 935
DB 452 AATTGGAGCTTAAGCCCAAGGAAGTGTCTTGGACCCCTGAGCAAGAGCTTTCTCTTAG 511
QY 936 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCTCTCAATCCCGCTGTGACTCTA 995
DB 512 AGCAGCAGCTGTGGGAGTGGAGCTGTGGCCAGTAGCTCTCAATCCCGCTGTGACTCTA 571
QY 996 ACCGGGAACCTTGGCTCTTCGCCCTCTGGGCCCTCTAAACAATCAACGCATAGACAATTTCA 1055
DB 572 ACCGGGAACCTTGGCTCTTCGCCCTCTGGGCCCTCTAAACAATCAACGCATAGACAATTTCA 631
QY 1056 GTGTGGTATCTATTGTCTGAACCCCTTCTACCTGGACTGCTTCTTAGGCTAGAGTCA 1115
DB 632 GTGTGGTATCTATTGTCTGAACCCCTTCTACCTGGACTGCTTCTTAGGCTAGAGTCA 691
QY 1116 GCTATCAGAACAGTCAAGTAATACGAGGTAGTCGGCTTCAGACACTAACCCCAAGAGA 1175
DB 692 GCTATCAGAACAGTCAAGTAATACGAGGTAGTCGGCTTCAGACACTAACCCCAAGAGA 751
QY 1176 GCAGCTGCCTGAGAGAGGACCAATTGTTCATCCCCAGCCGCTGAGTGCAGAACTCATTTCCGG 1235

DB 752 GCAGCTGCCTGAGAGAGGACCAATTTGTTCATCCCCAGCCGTTGARTGCAGAACTCATTCGG 811
QY 1236 CTTCTGTGGCAGGATGTCACCTCTTTCTACGGAAGSCCTTACCAAGAAATTCACCATCTTC 1295
DB 812 SCTGTGGCAGGATGTCACCTCTTTCTACGGAAGSCCTTACCAAGAAATTCACCATCTTC 871
QY 1296 GCATGAACCGGCTGGAAATCTCTTCAACAGGCTAACAGGGGCAAGATTATCCCAACCCCTG 1355
DB 872 GCATGAACCGGCTGGAAATCTCTTCAACAGGCTAACAGGGGCAAGATTATCCCAACCCCTG 931
QY 1356 ACCAGGATATGCTACCAAGGCTGAGGAGGAACAACAGCTTCTCCGATGATGATCAAA 1415
DB 932 ACCAGGATATGCTACCAAGGCTGAGGAGGAACAACAGCTTCTCCGATGATGATCAAA 991
QY 1416 AACACTCAGAGATTAACCCCAACAGCTTCTCTGCTGCTGAGACATCTCTCTGGAACA 1475
DB 992 AACACTCAGAGATTAACCCCAACAGCTTCTCTGCTGCTGAGACATCTCTCTGGAACA 1051
QY 1476 CCCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTCCACTTGTCTTGG 1535
DB 1052 CCCAGGAATCCACTGAAGAAAAAATAGAAATTAATACTACAGAGGTTCCACTTGTCTTGG 1111
QY 1536 AAGAAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGATACCTATGGAAGAGAGCCTG 1595
DB 1112 AAGAAGAGAGCCCTTCTGAGGGCTGTCCATCTAGTGAGATACCTATGGAAGAGAGCCTG 1171
QY 1596 GAGAGGCCCAAGTAAGTGATGTTGATTTACTCATCTAGAGGTGACCTTCCCATTTCTG 1655
DB 1172 GAGAGGCCCAAGTAAGTGATGTTGATTTACTCATCTAGAGGTGACCTTCCCATTTCTG 1231
QY 1656 CCAGACCAGCTTGTAGTAAACAAACTGATAGATTATATTTTGGGAGGTGCATCCAGTGACC 1715
DB 1232 CCAGACCAGCTTGTAGTAAACAAACTGATAGATTATATTTTGGGAGGTGCATCCAGTGACC 1291
QY 1716 TGGAAAACAAGTTCTGATCCAGAGGTGAGATTGGGATGAGGAAGCTGAGGATGATGTT 1775
DB 1292 TGGAAAACAAGTTCTGATCCAGAGGTGAGATTGGGATGAGGAAGCTGAGGATGATGTT 1351
QY 1776 TTGATAGTGTAGTCTACTGTGAGACTCAGACTTGAACCAAGACCCCTGAAGGGCTTCACC 1835
DB 1352 TTGATAGTGTAGTCTACTGTGAGACTCAGACTTGAACCAAGACCCCTGAAGGGCTTCACC 1411
QY 1836 TTTGGAACTCTTTCTGCGAGTGTAGATCCTTATAATCCCAAGAACTTTTACAGCAACAATTC 1895
DB 1412 TTTGGAACTCTTTCTGCGAGTGTAGATCCTTATAATCCCAAGAACTTTTACAGCAACAATTC 1471
QY 1896 AGACTGCTGCCAGAAATTTGTTCTTGAAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1955
DB 1472 AGACTGCTGCCAGAAATTTGTTCTTGAAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1531
QY 1956 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 2015
DB 1532 AGTCTGATCTAGAGAAATTTCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 1591
QY 2016 CTGGGAGGAGAGATGACTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGA 2075
DB 1592 CTGGGAGGAGAGATGACTGGGAATCTAGTCAGATGAAGCAGAGAGTCTCAAACTGTGGA 1651
QY 2076 ACTCATTTCTGTAATTCGTATGACCCCTTCAACCCCTTTAAATTTTAAAGGCTCTTTTCAA 2135
DB 1652 ACTCATTTCTGTAATTCGTATGACCCCTTCAACCCCTTTAAATTTTAAAGGCTCTTTTCAA 1711
QY 2136 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTCTGG 2195
DB 1712 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTGG 1771
QY 2196 CCATTTCTGAGTGTCAACCTTACTTTTGTAAAGTGTGAGTGTGGGAGGAGCAAGAA 2255
DB 1772 CCATTTCTGAGTGTCAACCTTACTTTTGTAAAGTGTGAGTGTGGGAGGAGCAAGAA 1831
QY 2256 GTGATGCTCAGACTCGGTACAGGGTGACCTTTCTTCTGGAGGAGACACACATCTCA 2315

Db 1832 GTGAATGTCAGACTCGGTGACGCGTGAGCTCTTTCTTGGAGGAGACACACATGTCA 1891

Qy 2316 AAAGAAAAAGTAACTCTCTTGAAGAAGTTACTAGTATTATATAGTGGTATGAGG 2375

Db 1892 AAAGAAAAAGTAACTCTCTTGAAGAAGTTACTAGTATTATATAGTGGTATGAGG 1951

Qy 2376 ATCGCAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGAGTTCAGAAAACGAATTC 2435

Db 1952 ATCGCAAGGACCATGGGAAGAAATTTGCAAGGGATGGATGAGTTCAGAAAACGAATTC 2011

Qy 2436 AAGAAACAGAGATGCTATTGCGATATTCCTTGACATTTGAACACAGAGAAAGATGTTA 2495

Db 2012 AAGAAACAGAGATGCTATTGCGATATTCCTTGACATTTGAACACAGAGAAAGATGTTA 2071

Qy 2496 ATAGACTCCAGGGAACATGCTTCAAGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCA 2555

Db 2072 ATAGACTCCAGGGAACATGCTTCAAGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCA 2131

Qy 2556 GCCTGTAGTCTAGTACATACACTACTCTTACCTGAGAGGTGTCTTTTAAAAACAAA 2615

Db 2132 GCCTGTAGTCTAGTACATACACTACTCTTACCTGAGAGGTGTCTTTTAAAAACAAA 2191

Qy 2616 TCTTGGCAGCTGCTCTTGGACATTTTCTTTTATAGAGGAATGTAACTTGGATCTAGTTT 2675

Db 2192 TCTTGGCAGCTGCTCTTGGACATTTTCTTTTATAGAGGAATGTAACT--GGMCTGTAA 2247

Qy 2676 AATTTTTTTTTTGCACATATCCCACTCAGAAACATTT--CAGGTTTGAAGCCAGCCCTT-G 2733

Db 2248 TTTTTTTTTTGCACATATCCCACTCAGAAACATTTCCAGGTTTGAAGCCAGCCCTTG 2307

Qy 2734 ATAATGAAGGATGAAC 2750

Db 2308 ATAATGAAGGATGAAC 2324

RESULT 4

ADQ88459

ID ADQ88459 standard; cDNA; 5468 BP.

XX AC ADQ88459;

XX XX

DT 07-OCT-2004 (first entry)

XX XX

Mouse GADD34-like (GADD34L) cDNA.

DE Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;

XX autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;

KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; mouse; gene;

XX ss.

XX OS

XX Mus sp.

XX Key

XX Location/Qualifiers

FT CDS 462..2558

FT /*tag= a

FT /product= "Mouse GADD34-like (GADD34L) protein"

FT /transl_except= (pos:1332..1334, aa:Arg)

FT /transl_except= (pos:1341..1343, aa:Arg)

FT /transl_except= (pos:1344..1346, aa:Trp)

FT /transl_except= (pos:1350..1352, aa:Val)

FT /transl_except= (pos:1362..1364, aa:Pro)

FT /transl_except= (pos:1368..1370, aa:Gln)

FT /transl_except= (pos:1377..1379, aa:Asp)

FT /transl_except= (pos:1389..1391, aa:Leu)

XX XX

PN US2004142345-A1.

XX XX

PD 22-JUL-2004.

XX XX

PF 28-AUG-2003; 2003US-00650482.

XX XX

PR 06-SEP-2002; 2002US-0408679P.

XX XX

PA (ROND/) RON D.

PA (JOUS/) JOUSSE C.

PI Ron D, Jousse C;

XX WPI; 2004-552556/53.

DR P-PSDB; ADQ88460.

XX Screening test substances for preventing or treating disease involving oxidative stress, by testing test substances for its ability to inhibit activity of GADD34L and identifying test substance that inhibits activity of GADD34L.

PS Disclosure; SEQ ID NO 3; 30pp; English.

XX The present invention relates to a method of screening several test substances for preventing or treating diseases involving oxidative stress such as neuronal ischaemia, heart ischaemia, renal damage induced by ischaemia or toxins, autoimmune diseases and neurodegenerative disorders. The method involves testing the test substances for its ability to inhibit the activity of GADD34-like (GADD34L), also referred to as eIF2alpha-specific regulatory subunit of phosphatase, and identifying the test substance which inhibits the activity of GADD34L. The present sequence is mouse GADD34L cDNA.

XX Sequence 5468 BP; 1321 A; 1272 C; 1366 G; 1509 T; 0 U; 0 Other;

Qy Query Match 45.0%; Score 1324; DB 12; Length 5468;

Db Best Local Similarity 72.5%; Pred. No. 0;

Qy Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;

Qy 96 CCATCGCCGAGGCGGTGGGGGACTCTCATTCATGGTGTGAAAGCGTCGAGCCGACTAGG 155

Db 151 CCGTCGCGCGCGCGAGGAGGGTCTTCTTATGGTGGAGCATCTCACAGGCGCTAGG 210

Qy 156 GAACCTCTTCCC-CGCCAGGATGGAAGTGCATCAGTCGCCGCTATTTCGGCGGGCTGT 214

Db 211 ACGTCTCTCTTCCCTAGCGGGATGGACCTAAACCGCGTCCGCCCTTGGCGGGCGCTC 270

Qy 215 TCTTCCCTGTGTCTGCGCGCGCTGCGCATTCGCTGCGCTCTGTGGCTTTTCTGTGGT 274

Db 271 TGGGCGCTCGGTGCAGCACTCGTTGCGGAAGCGCGCTCTCTGGGCTCTCTCTGCGG 330

Qy 275 CTCGAAGATCGGCTTGAGCAGCAGCCACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334

Db 331 CGCGGAATCGGACTGCAGTACCACTCCGTGGCTGGGCAAGCGCGAGACTGTGTAGACC 390

Qy 335 TCCTCGAATCCCGTGCACCTCAGCGCTGAGCGCGCGCGGCTTACCTGAGAGACTGTC 394

Db 391 TCGGATCCAGCTCGCTGACGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAAGCGGCC 450

Qy 395 AAGAAAAGAGATGGAGCGGCGGACAGCGGATCGCGGAAACGGCTTGGCCCTCGGGCG 454

Db 451 AAG-GAAGAGATGGAGACAGGAAACGACAGGCGCGGAGCGGCTTGGCCCTCGGCTG 509

Qy 455 GGCTTCGGTCTTGGCCACCCCTTTTTCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTC 514

Db 510 GGCTCTCTGTTCCGGCTGCCC---TTCCTTCGGGATCGCACGCTCTCTCGGAGTTC 566

Qy 515 CGAGCGCTCTTGGCGCGGAAACTCGGGAAACCCACACTGCTTCTCTGCCAGCCCC 574

Db 567 CCGCGCGCTTCTCTCGACAAAATCCCGGAA-----CTCCGCTCTGCCCC 611

Qy 575 GAGACTCGGTGAGTTACTGCGAAGAACTGCTCTCCAGCTCTTGGCGCGCTCCCCGGA 634

Db 612 GAGGTCGGAACAGGATCTGACCAAAATGCTTCTTCTCAGCTCTTGGCCCTGCTCCCTAGC 671

Qy 635 TTGCTTCAGAAGGTGCTAATTTTGGAGCCAACTTTTGGTGGAAATGTTTCCGACAGATGG 694

Db 672 CTATTCAGAAGCTGCTGCTTTTGGAGCCAGCTTTTCCGGGGCGCTGATTCTTACAGATGG 731

Qy 695 CTAGATTTTGTGAGTCTATACAGGCGGCTGAGAGCCCTGAGGAGCGGGAACCAACGACC 754

Db	732	CTAGATTTTGGCGCAAGTTTACAGCGCCTCTGAGAGCTTCGAGAGACCGGAGGAAATCTGCAC	731
Qy	755	GCCCCCAAGCGAGAGAAATCTTTGAGTTT-CGCTCGAGCTCGACTCTCTC---AGACCCCTC	810
Db	792	GCTCCACGGTGCAGAGTCTCTCAGTTTACATCGCGCTGAGCTCTTTCGCGAAGACTCGC	851
Qy	811	GGTCACAGTCCCTTGATTTGGCTTAGAGGAGGGATCCATGCGCAATACTCTGCGCCCCAGA	870
Db	852	GTCTCAGTACTCTTTGCAATTTGGCTTAGAGGAGGGACTCCAGTGGCAGTGTCTGCTCTCAGA	911
Qy	871	CCTAAAAATTGAGAGCTTAAGGCCAAAGGAAGTGTCTTTGGAACCTTCGACGACAGGCTTTTCT	930
Db	912	-CTGGAAGTTAAACTCAAGGCCCGAGGAAGAGCTTTAGACTCTCTGACGCGCCACTTTCTCT	970
Qy	931	CTTAGAGCAGCAGCTGTGCGGAGTGGAGACTGTCTGCCAGTAGCCTTCAATCCCGTCTGTGA	990
Db	971	CCTGGAGCAGCAGCTGTGGGAGTGGAGTTGCTGCCCAGTAGCCTTCAAGCTGTGTCTAGT	1030
Qy	991	CTCTAAACGGGAACTTGCTCTTCGCCCTCTGGCCCTCTAAACATTCAAACGATAGACAA	1050
Db	1031	CTCCACCGAGAACTTGACTCTTTCATCTCTGGGCCCTCTGAGCGTTTCAGAGCTTAGGTAA	1090
Qy	1051	TTTCAGTGTGTATCTCTATTTGTGTGAACCCCTTCTTACTCTGSACTGTCTTCTTAGGCTAGA	1110
Db	1091	TTTCAAGGTAGTTTCTCTATCTCTGAAACCCCTTCTCTCTGAGCTTACCTTCCCAAGTTAGG	1150
Qy	1111	AGTCAGCTATCAGAAACAGTGTAGTAATAGCGAGTGTGGCTTCAGACACTTAACCCC	1170
Db	1151	GCTCGCTGTCTCAGACAGCGCTGAGGTGGCCAGTTTGTGGTTCGGAACACTTAACCCC	1210
Qy	1171	AGAGCAGCTGCCTCTGAGAGAGGACCAATTGTCTATCCCGAGCGCTGAGTGCAGAACTCAT	1230
Db	1211	AGAGAGTGTCTATCTTTCTGAAGATGTTGTCAACCTCAGCGCTTGGCGGACAGATGTC	1270
Qy	1231	TCCGGCTCTGTGGCAGGATGTCCACCTCTTTCTACGGAAGGCCCTACAGAAATTCACCA	1290
Db	1271	GGCAACCGCCTGGAGAAGGTGTCCGCCCTCTCTCTACAGAAGGCCGTGCGGAAATTCACCA	1330
Qy	1291	TCTTTCGATGAACCGCTGGAAATTCCTTTCAACAGGCTTAAACAAGGGCAAGATTTACCCAC	1350
Db	1331	CTTTCTGTATGAAACGGCTAGAAATTCCT---CCAGCTTAAACAGGGCAAGTTTACCCAC	1387
Qy	1351	CCCTGACCAAGTAATGGCTACCAACAGCCTGGAGGAGGAACACAGCCTTCTCCGATGGA	1410
Db	1388	CCCTGACCAAGATAATGGCTATCATAGCTGGAGGAGGAACATAACCTTCTCCGATGGA	1447
Qy	1411	TCCAAACACTGCAGAGTAACCCAAACACAGTTTGTTCCTGTCTGTCTGGAGACATTCCTGG	1470
Db	1448	CCCAACAATTGCAACATACCCAGCACAGGCGGTTGTCCCTGTCTGCACAGAGGC---	1504
Qy	1471	AAACACCCAGGAATTCCTGAAGAAAAAATAGAAATTAATACTACAGAGTTCACATTGC	1530
Db	1505	-----GGAGCCCACTGAGAAAAAACCAAGATTGTTGATTCAAGAGTTTCAC-----	1551
Qy	1531	TTTGGAGAAGAGAGGCCCTTCTGAGGCTGTCCATCTAGTGTAGATACCTATGAAAAAGGA	1590
Db	1552	-----AGAGCCCCCAGGGAAGCAGTGTGTTTTGTGAATATACCGTGGAAAAAGA	1600
Qy	1591	GCCTGGAGAGGGCCGAATAGTGTAGTTGATTACTCATACTAGAGGTCACCTTCCCAT	1650
Db	1601	ATGTGAAGAGGACCACTAATGCACCTGACCTCTCAGATAGAGGAGAGCCCTTCTGT	1660
Qy	1651	TTCTGCCAGACCAAGCTTGTAGTAACAAACTGATAGATTATTTTGGAGGTCATCCAG	1710
Db	1661	TTCTACCAAGCCAGTTTGTAGCAACAACCTGATAGATTATTTTGGAGGCGCCCCCAG	1720
Qy	1711	TGACCTCGAAACAGTTCTGATCCAGAAGGTGAGGATTTGGGATCAGGAACCTCAGGATGA	1770
Db	1721	TGACTTGGAGCCAGCTCTGATCTGAAAGTAGGATTTGGGGCAGGAACCTCAGAGACGA	1780
Qy	1771	TGGTTTTGTAGTGTATAGCTCAGTGTCTGAGACTCAGACTCTTTGAAACAAGACCTTGAAGGGCT	1830
Db	1781	TGGCTTTGTATAGCAATGGCTCCCTGTCTGAATCAGAGCTGGAACAGGACTCGGAAGGCCT	1840

Qy	1831	TCACCTTTGGAACTCTTTCTGCGAGTGTAGATCCTTATATCCCCAGAACTTTTACAGCAAC	1890
Db	1841	TCACCTTTGGAACTCTTTTCCACAGTGTAGATCCTTACAAACCCCAAACTTTTACAGCCAC	1900
Qy	1891	AAATTACAGCTGCTGCCAGAAATTGTTCTGAAGAGCCTTCTGTATTCAGAGAAGATTTGTC	1950
Db	1901	GATTCAGACGCTGCCAGAAATTCGCCCCACAGAGACCCATCAGATTCAGGGACATCCTGGTC	1960
Qy	1951	TGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAAGCCTTCCTGAGACCCCTGAGCA	2010
Db	1961	TGGCAGCTGTGGTGTAGGGA---GCTGTCCAGGAGGACCCCTTCGAGAGACCCCGACCA	2017
Qy	2011	TAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTCAGAGATGAAGCAGAGAGTCTCAAACT	2070
Db	2018	TAGTTCGGGGAGGAAGATGACTGGGAACCGAGTCAGATGAAGCAGAGAAATCTTAAATT	2077
Qy	2071	GTGGAACTCATTTCTGTAAATTCGATGACCCCTCAACCCCTTTAAATTTTAAAGCTCCTTT	2130
Db	2078	GTGGAACTCTTTCTGTCAATTCAGAGACCCCTCAAACTTTTAAATTTTAAAGCTCCTTT	2137
Qy	2131	TCAAAACATCAGGGGAAAAATGAGAAAGGCTGTCTGTACTCAAAAGACCCCATCTGTAGTCCAT	2190
Db	2138	TCAACCGTCAGGGAAGAAATTGGAAGGCCGTGAGHACTCAAGGCCCTCTCTGAGGTCAAC	2197
Qy	2191	TGTGGCCATTTCTGAGTGTCAACACTTACTTTCTGTAAAGTGTGACGCTGTTGGGAGCCA	2250
Db	2198	AGTGGCCCTTCTCTGGCCATCATACCTTACTTTCTTGTAAAGGCCACGCTGTTAGAGAGCCA	2257
Qy	2251	AGAAAGTAATGTCACAGCTCGGTACAGCGTGACGTTCTTTCTGGAGGAAGACACACACA	2310
Db	2258	AGAAGATAATTGTCAGGCTGTGGGTGGGTGAGGCTCTTGCTGGGAAGAAGATACACCCA	2317
Qy	2311	TGTCAAAAGAAAAAGGTAACTTCTCTGGAAGAAGTTACTGAGTATTTATTAAGTGGTGA	2370
Db	2318	TATCAAGAGAAAAAGGTAACTTCTCTGGAAGAAGTTACTGAGTATTTATTAAGTGGTGA	2377
Qy	2371	TGAGGATCGCAAAAGGACCATGGGAAGAAATTGCAAGGGATGGAATGCAGGTTCCAGAAAAACG	2430
Db	2378	TGAGGATCGCAAGGACCATGGGAAGAAATTGCAAGGGATGGAATGCAGGTTCCAGAAAAACG	2437
Qy	2431	AATTCAGAAACAGAGATGCTATTTGGATATTGCTTTGACATTTGTGAACACAGAGAAAGAAT	2490
Db	2438	AAATTCAGAAACAGAGATTTGCCATTTGGCTACTGCTTTGGCCTTTTGACACAGAGAAAAAAT	2497
Qy	2491	GTTTAATAGACTCCAGGGAACATGCTTCAAAGACATTAATGTTCTCAAGCAATGTTAGT	2550
Db	2498	GTTTAATAGACT----GAGATCAGTCAAGACATTAATGTTGTACAGCAATGTTAAGA	2553
Qy	2551	TGCAGCCTGTAGTCTCTAGCATACACTACCTCTTACCTGAGAGGTGCTCTTTTAAAA	2610
Db	2554	AGTGA-ACAGCCTGCAACCGTGCCCACTCTGTCTTACTTTGAGA-GTTTCCCTTTAAAA	2611
Qy	2611	ACAAATCTTGGCAGCTGCTCTTGACATTTTTTTTTTTTATAGAGGAATGTAACTTGGATCT	2670
Db	2612	ACAAACACTGGCAGCTGCTCTTGACATGTTTTTAAAGAAACAACTTGTATCTAGAGATG	2671
Qy	2671	AGTTTAATTTTTTTTTTGCAACATATCCCACTCAGAAACATTCAGGTTTGAAGCCAGCC	2730
Db	2672	CAGTTTGAATATTTTGGGTAAATGTCTCATTAGAA-----ACACCAACT	2717
Qy	2731	CTCATATGAAGGATGAACCTAGTGATGATTTCTTAATCCTCCCTTTTGTATTTAGTTGGAT	2790
Db	2718	CCGATATGAAGAAAT-----CTCTTATCTGTAAATCCTCTCTTTTCCATATTTAGTTGGAT	2771
Qy	2791	GTGCTTTT	2798
Db	2772	GTGGGTTT	2779

RESULT 5
AAH99389
ID AAH99389 standard: cDNA: 883 BP.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00300253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00183767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX PA (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX Claim 1; SEQ ID NO 3458; 2537pp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 XX Sequence 772 BP; 134 A; 256 C; 217 G; 161 T; 0 U; 4 Other;
 SQ
 Query Match 25.1%; Score 738; DB 4; Length 772;
 Best Local Similarity 98.8%; Pred. No. 5.7e-209; Indels 1; Gaps 1;
 Matches 752; Conservative 0; Mismatches 8;
 QY 1 ATTTTGGGCTTCGTTCCACCGCAGCGGCTTACCGAGTCTTCCGGTATCGCGTTG 60
 DB 1 ATTTTGGGCTTCGTTCCACCGCAGCGGCTTACCGAGTCTTCCGGTATCGCGTTG 60
 QY 61 CTCAGGGGCTTTTCAACCTCTCTCAGTCGGAAACCATCGCGAGCGGCTGGGGGACT 120
 DB 61 CTCAGGGGCTTTTCAACCTCTCTCAGTCGGAAACCATCGCGAGCGGCTGGGGGACT 120
 QY 121 CCTATCCATGGTGTGAAGCGTCGAGCGGCTAGGGAACCTCTTCCCGCGCAGGATGGA 180
 DB 121 CCTATCCATGGTGTGAAGCGTCGAGCGGCTAGGGAACCTCTTCCCGCGCAGGATGGA 180
 QY 181 AGTCGCATCAGTCGCGCGCTATTCGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 240

DB 181 AGTCGCATCAGTCGCGCGCTATTCGCGGGGCTGTTCTTCCCTGTGTTCTGCGCGCGCTG 240
 QY 241 CCGCATTTGGCTGCGCTCTGTGGCTTTTCTGCTGGCTCGAGATCGGCTGAGCAGCGAC 300
 DB 241 CCGCATTTGGCTGCGCTCTGTGGCTTTTCTGCTGGCTCGAGATCGGCTGAGCAGCGAC 300
 QY 301 GCCACCGCTGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360
 DB 301 GCCACCGCTGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCCAGCTCCAGC 360
 QY 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCGGGGAC 420
 DB 361 CGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCGGGGAC 420
 QY 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGTTCCTGCGCCACCTTTT 480
 DB 421 AGCGGATCGCGGAAACGGCTTGGCCCTCGGGGGGCTTCGGTTCCTGCGCCACCTTTT 480
 QY 481 CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTTCCCGACGCTCTTGGCCCGGAAACTC 540
 DB 481 CCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTTCCCGACGCTCTTGGCCCGGAAACTC 540
 QY 541 CGGAAACCCACACTGTTCTTCTGCGCGGCGGAGACTCGGGTCAAGTCTAGTACGAGAA 600
 DB 541 CGGAAACCCACACTGTTCTTCTGCGCGGCGGAGACTCGGGTCAAGTCTAGTACGAGAA 600
 QY 601 ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGTTTCAGAAAGTGTCTAATTTGAG 660
 DB 601 ACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGTTTCAGAAAGTGTCTAATTTGAG 660
 QY 661 CCAACTTTTTCGTTGGAATGTTTCCGACGAGATGGCTAGATTTTCTGAGTCTAC-AGCG 719
 DB 661 CCAACTTTTTCGTTGGAATGTTTCCGACGAGATGGCTAGATTTTCTGAGTCTAC-AGCG 720
 QY 720 CCCTGAGAGCGCTGAAGGAGCGGAGAAACAGCGCGGCGCC 760
 DB 721 CCCTGAGAGCGCTGAAGGAGCGGAGAAACAGCGCGGCGCC 761
 RESULT 7
 AAH11666/c
 ID AAH11666 standard; cDNA; 543 BP.
 XX AC AAH11666;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human cDNA clone (3'-primer) SEQ ID NO:8501.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX OS Homo sapiens.
 XX FN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00300253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00183767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX PA (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 3; SEQ ID NO 8501; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 543 BP; 189 A; 104 C; 97 G; 149 T; 0 U; 4 Other;

Query Match 17.5%; Score 515; DB 4; Length 543;
 Best Local Similarity 96.5%; Pred. No. 1.8e-142;
 Matches 524; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

2400 TTGCAAGGATGGATGAGTTCAGAAACGAAATTCAGAAACAGAGATGCTATTGGAT 2459
 |||||
 543 TTTTCAAGGATGATCCAGTTCNAGAAAGGATTCAGAAACAGAGATGCTATTGGAT 484
 |||||

2460 ATTGCTTGACATTTGAACACAGAGAAAGATGTTTAAATAGACTCCAGGACATGCTTCA 2519
 |||||

483 ATGCTTGACATTTGACCCACAGAGAAAGATGTTTAAATAGACTCCAGGACATGCTTCA 424
 |||||

2520 AAGGACTTAATGTTCTCAAGCAATGTTGAGTTGCGAGCTGTAGTCTAGCTAGCATACA 2579
 |||||

423 AAGGACTTAATGTTCTCAAGCAATGTTGAGTTGCGAGCTGTAGTCTAGCTAGCATACA 364
 |||||

2580 CTACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATT 2639
 |||||

363 CTACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATT 304
 |||||

2640 TTTTCTTTTACAGGAAATGTAATGATCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 2699
 |||||

303 TTTTCTTTTACAGGAAATGTAATGATCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 244
 |||||

2700 CACTCAGAAACATTCAGGTTTGAAGCCAGCCCTCATATATGAGGATGAATGATGATGATT 2759
 |||||

243 CACTCAGAAACATTCAGGTTTGAAGCCAGCCCTCATATATGAGGATGAATGATGATGATT 184
 |||||

2760 TCTAATCTCTCTCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2819
 |||||

183 TCTAATCTCTCTCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 124
 |||||

2820 GGTGAAGAGGAGCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2879
 |||||

123 GGTGAAGAGGAGCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 64
 |||||

2880 AACATATTTATAGGCTTTAAGCCCACTTTTCAATCTTCAATCTTCAATCTTCAATCTT 2939
 |||||

63 AACATATTTATAGGCTTTAAGCCCACTTTTCAATCTTCAATCTTCAATCTTCAATCTT 4
 |||||

2940 CTG 2942
 |||||

Db 3 CTG 1

RESULT 8

ABL38451
 ID ABL38451 standard; cDNA; 561 BP.

XX ABL38451;
 AC ABL38451;
 DT 08-APR-2002 (first entry)

XX Human colon tumour antigen polynucleotide SEQ ID NO:2040.

XX Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW Human tumour metastatic antigen; diagnosis; gene; ss.

XX Homo sapiens.

XX WO200196388-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US018557.

XX 09-JUN-2000; 2000US-0210899P.

XX 20-FEB-2001; 2001US-0270216P.

XX (CORI-) CORIXA CORP.

XX Jiang Y, Harlocker SL, Secrist H;

XX WPI; 2002-114514/15.

XX Novel isolated colon tumor polynucleotide differentially expressed in colon tumor or colon metastatic tumor and polypeptides encoded by them, useful for inhibiting development of cancer in patient.

XX Claim 1; SEQ ID NO 2040; 105pp; English.

XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I) which were isolated from human colon tumour and colon metastatic tumour cDNA libraries. (I) have cytostatic activity and can be used in vaccine production. (I) can be used for stimulating and/or expanding T cells specific for a tumour protein on contact with the T cells. They are also useful for inhibiting the development of cancer in a patient. (I) can be used as probes or primers for nucleic acid hybridisation, for preparing mutant species primers, or primers for use in genetic constructions. (I) can be used in the diagnosis of a colon tumour

XX Sequence 561 BP; 144 A; 91 C; 93 G; 226 T; 0 U; 7 Other;

Query Match 9.0%; Score 265.8; DB 6; Length 561;
 Best Local Similarity 96.4%; Pred. No. 5.1e-68;
 Matches 270; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2663 TTGGATCTAGTCTTAAATTTTTTTTTTGGCAACATATCCACTCAGAAACATTCAGGTTTGA 2722
 |||||

Db 3 TTGGNNCTAGTCTTAAATTTTTTTTTTGGCAACATATCCACTCAGAAACATTCAGGTTTGA 62
 |||||

Qy 2723 AGCCAGCCCTGATAAAGAGGATGAATCTAGTGTGATTTCTAATCTCTCTCTTTTGAATTT 2782
 |||||

Db 63 AGCCAGCCCTGATAAAGAGGATGAATCTAGTGTGATTTCTAATCTCTCTCTCTTTTGAATTT 122
 |||||

Qy 2783 AGTTGGATGTCTTTTAAATGTCTTTGCTGATGAGTGGTGAAGAGGACCTTTTGTAG 2842
 |||||

Db 123 AGTTGGATGTCTTTTAAATGTCTTTGCTGATGAGTGGTGAAGAGGACCTTTTGTAG 182
 |||||

Qy 2843 TTGTCTTTTGGCACTTTCAAAACCTTATTTCTTGGAAACAAATATTTATAGGCTTAAAG 2902
 |||||

Db 183 TTGTCTTTTGGCACTTTCAAAACCTTATTTCTTGGAAACAAATATTTATAGGCTTAAAG 242
 |||||

Qy 2903 CCCATTTTCTAATCTTAAATATTTGTGTGCTATCTG 2942
 |||||

Db 517 GCGTTAAAGCCCATTTTCATTTCTAAATTAATGTCGCTATCTG 469

RESULT 11
ABX52167
ID ABX52167 standard; cDNA; 399 BP.

XX AC ABX52167;
XX DT 25-FEB-2003 (first entry)

XX DE Bovine EST associated with lactation/muscle/fat deposition #2096.

XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.

XX OS Bos Taurus.

XX PN US2002137160-A1.

XX PD 26-SEP-2002.

XX PF 26-OCT-2001; 2001US-00983965.

XX PR 17-DEC-1998; 98US-0113678P.

XX PR 15-DEC-1999; 99US-00465231.

XX PA (BYAT//) BYATT J C.

XX PA (MATH//) MATHIALAGAN N.

XX PA (TAON//) TAO N.

XX PA (WARR//) WARREN W C.

XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX DR WPI; 2003-102386/09.

XX PS Claim 2; SEQ ID NO 2096; 38pp; English.

XX CC The invention relates to a purified nucleic acid molecule associated with
XX CC lactation or muscle and fat deposition (designated LMFD), derived from
XX CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
XX CC nucleic acid molecule comprising any of 5912 nucleotide sequences,
XX CC appearing as ABX50072-ABX55983, or complements of them. Also included are
XX CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
XX CC acid linked to a promoter and a 3' non-translated sequence that
XX CC functions in the cell to cause termination of transcription and addition
XX CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
XX CC (2) determining a level or pattern of a molecule in a bovine cell or
XX CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
XX CC of the 5912 nucleic acid sequences or its complement or fragment) with a
XX CC complementary nucleic acid molecule obtained from the bovine cell or
XX CC tissue, where hybridisation between the marker nucleic acid and the
XX CC complementary nucleic acid permits the detection of the molecule; and (b)
XX CC detecting the level or pattern of the complementary nucleic acid, where
XX CC the detection of the complementary nucleic acid is predictive of the
XX CC level or pattern of the molecule. The LMFD nucleic acid is used for
XX CC determining a level or pattern of a molecule in a bovine cell or tissue.
XX CC It is useful for genome mapping, gene identification and analysis, cattle
XX CC breeding, preparation of constructs for use in cattle gene expression, or
XX CC for genetically improving cattle. The present sequence is one of the 5912
XX CC bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present
XX CC sequence was not shown in the specification but was obtained in
XX CC electronic format from the USPTO web site:
XX CC seqdata.uspto.gov/sequence.html?docID=20020137160

XX SQ Sequence 399 BP; 133 A; 62 C; 82 G; 122 T; 0 U; 0 Other;

Query Match 6.8%; Score 198.6; DB 8; Length 399;
Best Local Similarity 83.8%; Pred. No. 4.8e-48;
Matches 249; Conservative 0; Mismatches 44; Indels 4; Gaps 2;
QY 2323 AAAGGTAACCTTCCTTGAAGAAGCTTACTGAGTATTATATAAGTGGTGATGAGGATCGCAA 2382
DB 102 ATAGGTAACCTTCCTTGAAGAAGTCTTACTGAGTATTATATAAGCGGTGATGAGGATCGAAA 161
QY 2383 AGGACCATGGGAAGAAATTTGCAAGGGATGGATCGAGTTCAGAGAAACGAATTCAGAAGAAC 2442
DB 162 AGGACCATGGGAAGAAATTTGCAAGGGATGGATCGAGTTCAGAGAAACGAATTCAGAAGAAC 221
QY 2443 AGAAGATGCTATTGGATATTGCTTGACATTTGACACACAGAGAAGAAATGTTTAAATAGA-- 2500
DB 222 AGAAGAAGCTATTGGATCTGCTTGACATTTGAGCACAGAGAGAAAAGATGTTTTAAATATAA 281
QY 2501 -CTCCAGGGGAACATGCTTCAAGGAGCTTAAATGTTCTCAAGCAATGTTGAGTTG-GCAGCC 2558
DB 282 CTTCAAGCAACATGTTTCAAGGAGCTTAAATGTTTTCGAGCAATGTTNAGATGATTCGAC 341
QY 2559 TGTAGTCTTAGCTAGCATACACATCTTACCTCTGAGAGGTGCTTTTAAAAACAAA 2615
DB 342 AGCCTCTGGCCCTAGCATACACTACCTCTTACTTGAGAGGGTACTTTTAAAAACCAA 398

RESULT 12

AAC06938
ID AAC06938 standard; cDNA; 195 BP.

AC AAC06938;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 11013.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX KW Gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-00200610.

XX PR 26-FEB-1999; 99US-0122487P.

XX PA (GEST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX CC New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX CC obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX CC diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 1; SEQ ID NO 11013; 71pp + Sequence Listing; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors
SQ Sequence 195 BP; 33 A; 65 C; 54 G; 43 T; 0 U; 0 Other;
Query Match 6.6%; Score 195; DB 3; Length 195;
Best Local Similarity 100.0%; Pred. No. 3.6e-47;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATTTTGGGCTTGGCTTCCACCGCAGCCGCGCTTACCCAGTCTTCCGGTATCGCGTTG 60
Db 1 ATTTTGGGCTTGGCTTCCACCGCAGCCGCGCTTACCCAGTCTTCCGGTATCGCGTTG 60
Qy 61 CTCAGGGGCTTTCAACCTCTGTGTCAGTCGGAAACCATCGCCGAGGCGGTGGGGGACT 120
Db 61 CTCAGGGGCTTTCAACCTCTGTGTCAGTCGGAAACCATCGCCGAGGCGGTGGGGGACT 120
Qy 121 CCTATCCATGTTGTTGAAGCGTCGAGCCGACTAGGGAACCTCTTCCCGCCAGGATGGA 180
Db 121 CCTATCCATGTTGTTGAAGCGTCGAGCCGACTAGGGAACCTCTTCCCGCCAGGATGGA 180
Qy 181 AGTCGCATCAGTCGC 195
Db 181 AGTCGCATCAGTCGC 195
RESULT 13
AAS61819
ID AAS61819 standard; cDNA; 628 BP.
AC AAS61819;
XX
XX 29-JAN-2002 (first entry)
XX Lung small cell carcinoma antigen, cDNA #360.
XX Human; cytostatic; antitumour; lung small cell cancer antigen; tumour;
XX lung cancer; ss.
XX Homo sapiens.
XX WO200177168-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-US011859.
XX 11-APR-2000; 2000US-0196780P.
XX 21-JUN-2000; 2000US-0213361P.
XX 01-SEP-2000; 2000US-0229763P.
XX 05-SEP-2000; 2000US-0230629P.
XX 14-SEP-2000; 2000US-0232565P.
XX 19-DEC-2000; 2000US-0257037P.
XX 08-JAN-2001; 2001US-0260796P.
XX (CORI-) CORIXA CORP.
XX Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX WPI; 2002-010896/01.
XX Lung tumor polynucleotide and polypeptides useful in therapy and
XX diagnosis of cancer especially lung cancer.
XX Claim 1; Page 256; 295pp; English.
XX The invention relates to novel isolated lung small cell cancer antigen
XX polynucleotides (I) and polypeptides (II) used in a method of detecting
XX cancer in a patient. The method is optionally performed by utilising
XX oligonucleotides (III), where the biological sample from the patient is
XX contacted with (III), detecting the amount of polynucleotide hybridised
XX to (III) in the sample and comparing the amount of polynucleotide to a
XX predetermined cut-off value and thereby determining cancer in a patient.
XX (I), (II) or antigen-presenting cells expressing (II) is useful for

CC stimulating and/or expanding T cells specific for a tumour protein. The
CC method comprises contacting T cells with one of the components under
CC conditions to permit the stimulation and/or expansion of the cells. A
CC composition comprising (I) is useful for stimulating an immune response
CC in a patient and for inhibiting the development of a cancer especially
CC lung cancer in a patient. An isolated T cell population is useful for
CC removing tumour cells from the biological sample and for inhibiting the
CC development of cancer in a patient. AAS61460-AAS61874 represent novel
CC human lung small cell cancer antigen coding sequences of the invention
XX
SQ Sequence 628 BP; 161 A; 101 C; 103 G; 263 T; 0 U; 0 Other;
Query Match 6.4%; Score 187.4; DB 6; Length 628;
Best Local Similarity 99.5%; Pred. No. 1.4e-44;
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2754 GTGATTTCTAATCCTCTCTTTTGGATTGATTTAGTTGGATGCTTTTAAATGCTCTTGCCT 2813
Db 1 GTGATTTCTAATCCTCTCTTTTGGATTGATTTAGTTGGATGCTTTTAAATGCTCTTGCCT 60
Qy 2814 GCATGAGTGGAAAGGGGACCTTTTGGAGTTGTCATTTTGCACCTTCCAAACTTATTTTC 2873
Db 61 GCTTGGAGTGGAAAGGGGACCTTTTGGAGTTGTCATTTTGCACCTTCCAAACTTATTTTC 120
Qy 2874 TTGGAACAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATGCTGT 2933
Db 121 TTGGAACAATATTTATAGGCTTAAAGCCCATTTTCAATTTCTAATCTAAATATGCTGT 180
Qy 2934 GCCTATCTG 2942
Db 181 GCCTATCTG 189
RESULT 14
ADD49398/c
ID ADD49398 standard; cDNA; 173 BP.
AC ADD49398;
XX
XX 15-JAN-2004 (first entry)
XX Human lung cancer associated cDNA 61483154.
XX Human; ss; lung cancer antigen; cytostatic; lung cancer; gene therapy;
XX vaccine; T-cell; tumour.
XX Homo sapiens.
XX OS
XX US2003194764-A1.
XX 16-OCT-2003.
XX 04-APR-2002; 2002US-00116712.
XX 05-APR-2001; 2001US-0282289P.
XX 05-OCT-2001; 2001US-0327511P.
XX (CORI-) CORIXA CORP.
XX Bangor CS, Switzer A;
XX WPI; 2003-844452/78.
XX New isolated polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cancer, particularly lung cancer.
XX Claim 1; SEQ ID NO 130; 250pp; English.
XX The invention relates to an isolated polynucleotide (a) comprising any of
XX the 666 fully defined nucleotide sequences appearing as ADD49269 -
XX ADD49395, ADD49396 and ADD49398, complements of (a); sequences of at
XX least 20 contiguous residues of (a); sequences that hybridise to (a)
XX under highly stringent conditions; sequences having at least 75 or 90%

identity to (a); or degenerate variants of (a). Also included are an isolated polypeptide (b) comprising: sequences encoded by the new polynucleotide; any of the 4 amino acid sequences fully defined in the specification; or sequences having at least 70 or 90% identity to the sequence in (a) or (b)), an expression vector comprising the above polynucleotide operably linked to an expression control sequence, a host cell transformed or transfected with the above expression vector, an isolated antibody, or its antigen-binding fragment, that specifically binds to the above polypeptide, an oligonucleotide that hybridizes to the above-mentioned nucleotide sequences under highly stringent conditions, a fusion protein comprising at least one polypeptide cited above, detecting the presence of a cancer in a patient (comprising: obtaining a biological sample from the patient; contacting the biological sample with a binding agent that binds to the polypeptide, or with the oligonucleotide cited above; detecting in the sample an amount of the polypeptide that binds to the binding agent, or an amount of a polynucleotide that hybridizes to the oligonucleotide); and comparing the amount of polypeptide, or polynucleotide that hybridizes to the oligonucleotide, to a predetermined cut-off value and then determining the presence of a cancer in the patient), a method for stimulating and/or expanding T-cells specific for a tumour protein (comprising contacting T-cells with the above polypeptide, polynucleotide or antigen-presenting cells that express the polynucleotide, under conditions and for a time sufficient to permit the stimulation and/or expansion of T-cells), an isolated T-cell population comprising T-cells prepared by the method, a composition comprising a first component selected from physiological carriers and immunostimulants, and a second component selected from the above polypeptide, polynucleotide, antibody, fusion protein, T-cell population and antigen-presenting cells that express the above polypeptide, stimulating an immune response in a patient (comprising administering to the patient the above composition) treating lung cancer in a patient (comprising administering to the patient the above composition and a diagnostic kit (comprising: at least one oligonucleotide cited above; or at least one antibody cited above and a detection reagent, where the detection reagent comprises a reporter group). The composition and methods are useful in diagnosing, preventing and treating cancer, particularly lung cancer. The present sequence is a lung cancer-associated antigen cDNA of the invention.

Sequence 173 BP; 73 A; 30 C; 28 G; 42 T; 0 U; 0 Other;

Query Match 5.5%; Score 162; DB 10; Length 173;
Best Local Similarity 99.4%; Pred. No. 2.4e-37;
Matches 173; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2625 CTGCTCTTGCACATTTTCTTTTATAGAGGAATGTAACCTGGATCTAGTTTATTTT 2684
173 CTGCTCTTGCACATTTTCTTTTATAGAGGAATGTAACCTGGATCTAGTTTATTTT 115
2685 TTTTGCAACATATCCCACTCAGAAACATTGAGGTTTGAAGCCAGCCCTGATGAAGGA 2744
114 TTTTGCAACATATCCCACTCAGAAACATTGAGGTTTGAAGCCAGCCCTGATGAAGGA 55
2745 TGAACATAGTGATTTCTAATCTCCCTTTTGTGATTTAGTTGATGCTTTT 2798
54 TGAACATAGTGATTTCTAATCTCCCTTTTGTGATTTAGTTGATGCTTTT 1

RESULT 15
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX ADA71938;
AC ADA71938;
XX ADA71938;
DT 20-NOV-2003 (first entry)
XX Rice gene, SEQ ID 5263.
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX Oryza sativa.
OS
XX

PN WO2003000898-A1.
XX 03-JAN-2003.
XX 22-JUN-2001; 2001WO-IB001105.
XX 22-JUN-2001; 2001WO-IB001105.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX Claim 27; SEQ ID NO 5263; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
SQ
Query Match 2.1%; Score 61.8; DB 8; Length 2000;
Best Local Similarity 11.7%; Pred. No. 1e-06;
Matches 110; Conservative 402; Mismatches 414; Indels 11; Gaps 3;
QY 1705 ATCCAGTACCTGGAAACAGTTCTGATCCAGAGGTGAGGATGGATGAGGAGCTGA 1764
DB 158 RKXSGSGWKTCTRRGARGSGWSSGAKYKSGSMKMMSSCGRSGCGRRSAYRYGT 217
QY 1765 GATGATGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1824
DB 218 SRXYGYKWTYYSASRCMEAYMTTSYACSSYVTCRKRKRMMKMRMSRSGW 277
QY 1825 AGGGCTTACCTTTGGAACTCTTTCTGCGAGTGTAGATCTTATATCCCGAGACTTTAC 1884
DB 278 YSMYKMMCTAYKK--SYSRWCYMYRGGWGRGATRYMGRGYMSRMAMMYKMYWYRG 334
QY 1885 AGCAACAATTCAGACTGCTGCCAGAAATTTCTCTGAGAGCCCTTCGATTCAGAGGAGGA 1944
DB 335 YGKMKRGMWAGRMMSMRMSKRWSS--KACYMYRWRWRWRWRWRWRWRWRWRWRWRWR 393
QY 1945 TTTGTCTGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCCTTCCTGAGACCC 2004
DB 394 KRXYKMRGYSRMRSCKRBRMMKRCRSGRAWKRCRGOMTCRMKSYGMWRKMSKRWASK 453
QY 2005 TGAGCATAGTTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2064
DB 454 YKMSRMWRKRRKKSRTTMMGKTGMMGTMCRCYKRSKMRKRCRRRRWRWRWRWRWR 513
QY 2065 CAACTGTGGAACATCTTCTGTAATTTCTGATGACCCCTCAACCCCTTTTAATTTTAAGGC 2124
DB 514 KRYMSARYMTMYCARKKYSYSAARARCRWYRGKGYWAGMMWRKRYKRYMTMYKMMWYKR 573
QY 2125 TCCTTTTCAACATCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2184
DB 574 KYSKSNYKMSYASCMKSKARAGAKMCKRKSMSKMSKMSKMSKMSKMSKMSKMSKMSK 633
QY 2185 GTCCATTTGGCCATTTCTTGAGTGTCACACTTACTTTCTTGTAAAGTGCGAGCTGTGGG 2244

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 12:27:30 ; Search time 439.026 Seconds
(without alignments)
10965.010 Million cell updates/sec

Title: US-10-650-482-1

Perfect score: 2942

Sequence: 1 attttggcttcctccac.....aaattatgtgctatctg 2942

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
 - 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
 - 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
 - 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
 - 5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	7.4	226	4	US-09-621-976-8778 Sequence 8778, Ap
2	218	7.4	218	4	US-09-513-999C-33878 Sequence 33878, A
3	195	6.6	195	4	US-09-513-999C-11013 Sequence 11013, A
4	51	1.7	7218	1	US-08-232-463-14 Sequence 14, Appl
5	45.4	1.5	1141	4	US-09-806-708B-22 Sequence 22, Appl
6	42.6	1.4	45365	4	US-09-949-016-11893 Sequence 11893, A
7	42.6	1.4	45456	4	US-09-949-016-17007 Sequence 17007, A
8	42	1.4	364	4	US-09-621-976-17202 Sequence 17202, A
9	41.4	1.4	399	4	US-09-806-708B-22 Sequence 22, Appl
10	40.8	1.4	1141	4	US-09-806-708B-22 Sequence 22, Appl
11	40.4	1.4	117807	4	US-09-949-016-15525 Sequence 15525, A
12	40.2	1.4	1468	4	US-09-949-016-3573 Sequence 3573, Ap
13	40.2	1.4	1560	4	US-09-949-016-1177 Sequence 1177, Ap
14	40.2	1.4	4252	4	US-09-919-039-334 Sequence 334, Appl
15	39.6	1.3	7218	1	US-08-232-463-14 Sequence 14, Appl
16	39.4	1.3	390416	4	US-09-949-016-16923 Sequence 16923, A
17	39.2	1.3	12526	4	US-09-949-016-13865 Sequence 13865, A
18	39.2	1.3	23439	4	US-08-956-171E-38 Sequence 38, Appl
19	39.2	1.3	23439	4	US-08-781-986A-38 Sequence 38, Appl
20	39.2	1.3	161607	4	US-09-949-016-12210 Sequence 12210, A
21	38.8	1.3	112705	4	US-09-949-016-15630 Sequence 15630, A
22	38.6	1.3	11979	4	US-09-949-016-12309 Sequence 12309, A
23	38.6	1.3	19980	4	US-09-949-016-13533 Sequence 13533, A
24	38.2	1.3	474	4	US-09-621-976-18033 Sequence 18033, A
25	38.2	1.3	90876	4	US-09-949-016-13271 Sequence 13271, A
26	38	1.3	7409	4	US-09-949-016-12689 Sequence 12689, A
27	38	1.3	7409	4	US-09-949-016-15978 Sequence 15978, A

Sequence 16908, A
Sequence 14611, A
Sequence 12680, A
Sequence 13084, A
Sequence 12211, A
Sequence 16509, A
Sequence 12900, A
Sequence 12412, A
Sequence 11028, A
Sequence 1224, Ap
Sequence 750, App
Sequence 1647, Ap
Sequence 16413, A
Sequence 16425, A
Sequence 1, Appl
Sequence 15545, A
Sequence 12286, A
Sequence 14711, A
Sequence 13379, A
Sequence 1597, Ap
Sequence 3435, Ap
Sequence 18717, A
Sequence 342, App
Sequence 16703, A
Sequence 14999, A
Sequence 15000, A
Sequence 13747, A
Sequence 7350, Ap
Sequence 16656, A
Sequence 17028, A
Sequence 23, Appl
Sequence 41, Appl
Sequence 36, Appl
Sequence 3119, Ap
Sequence 20, Appl
Sequence 68, Appl
Sequence 16, Appl
Sequence 5888, Ap
Sequence 27, Appl
Sequence 161, App
Sequence 22, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 41, Appl
Sequence 41, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1432, A
Sequence 18033, A
Sequence 2813, Ap
Sequence 14970, A
Sequence 25, Appl
Sequence 6809, A
Sequence 2813, Ap
Sequence 1, Appl
Sequence 74411, A
Sequence 74412, A
Sequence 204315, A
Sequence 35, Appl
Sequence 13915, A
Sequence 17244, A
Sequence 4637, Ap
Sequence 7, Appl
Sequence 5, Appl
Sequence 1242, Ap


```

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

```

```

Query Match 1.7%, Score 51; DB 1; Length 7218;
Best Local Similarity 3.7%; Pred. No. 0.00076;
Matches 12; Conservative 189; Mismatches 124; Indels 0; Gaps 0;

Qy 2228 AAGTGCAGCTGTGGGGAGCCAGAAAGTCAATGTCAGACTCGGTACAGCGTGCAGTT 2287
Db 1370 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1311

Qy 2288 CTTTCTGGAGAGACACACATGTCAAAGAAAGTAACTTCTTGTGAAGAGTT 2347
Db 1310 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1251

Qy 2348 ACTGAGTATTATATAAGTGTGATGAGGATCGCAAGACCATGGGAAGATTTGCAAG 2407
Db 1250 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1191

Qy 2408 GATGATGCAGGTTCCAGAAAGAAATTCAGAAACAGAGATGCTATTGGATATGCTTG 2467
Db 1190 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1131

Qy 2468 ACATTGTAACACAGAAAGAAATGTTTAATAGACTCCAGGAACATGCTTCAAGGACTT 2527
Db 1130 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1071

Qy 2528 AATGTTCTCAAGCAATGTTGATGG 2552
Db 1070 RRRATCGCAAGCTCCCTCGACCTG 1046

```

```

RESULT 5
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B

```

```

; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
; US-09-806-708B-22

```

```

Query Match 1.5%; Score 45.4; DB 4; Length 1141;
Best Local Similarity 11.6%; Pred. No. 0.01;
Matches 109; Conservative 325; Mismatches 498; Indels 9; Gaps 2;

Qy 1773 GTTTGATAGTAGTACTCTCTCAGACTCAGACCTTGAAAGAGCCCTTGAAGGGCTTC 1832
Db 1094 GTAWTTHAKRGATCMCYWYWTCTGTRRWCRTYAMRTWYTRSNANWSCATKBMWMTMKW 1035

Qy 1833 ACCTTGGAACTCTTTCTGCAGCTGATAGTCTTATATATATATATATATATATATAT 1892
Db 1034 ATKYRTAWYAMWCWNRNNNMWCATNGYAKSCATNNAMVATTTRWAAAYAAAKWAGNNM 975

Qy 1893 TTCAGACTGCTGCCAGAAATTTCTTCTGGAAGAGCCTTCTGATTCAGAGAGGATTGTCG 1952
Db 974 RMTGAAAGKMGCMAAWATMBWADTAGMCMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 915

Qy 1953 GCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCCTTCTGAGACCCCTGAGCATA 2012
Db 914 TACYNRAATNNKWAATHMKWTHGHSKRTRHHTCRRTKYNNNNNNNNNNNNNNNNNNNN 855

Qy 2013 GTTCTGGGAGGAAGATGACTGGGAATCTAGTCAGATGAGAGAGAGAGTCTCAAACTGT 2072
Db 854 RW-----MNAWMTTRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 802

Qy 2073 GGAACCTCATCTCTAATCTCTGATGACCCCTTACAACTTTTAAATTTTAAAGGCTCTTTC 2132
Db 801 CHYTTANABBCYRANNNNAARWARTCNNTFMHAAVTTTHTDWCYKTWMTWYDMMTMB 742

Qy 2133 AAACATCAGGGGAAATAGAGAAAGGCTGCTGACTCAAGACCCCATCTGAGTCCATTG 2192
Db 741 TTTTTRNMTTSTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 682

Qy 2193 TGCCCATTTCTGAGTGTACACCTTACTTCTTGTAGGTGAGCTGTTGGGAGGCAAG 2252
Db 681 VMERRRMTNTKRWYSTTRRHHYTGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 622

Qy 2253 AAAGTGAATGTCAGACTCGGTACAGCTGCTTCTTCTGAGGAGAACACACACATG 2312
Db 621 DGMVTRKKVKRRDTCCTYVDVWADSWWWYANMRCRDVTYTRNNTYCKSYAHSYHWSN 562

Qy 2313 TCAAAAGAAAAAAGTAACCTTCTTGAAGAACTTACTAGTATTATATAAGTGGTGATG 2372
Db 561 NAMWRYRSARNSSMARWTRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 502

Qy 2373 AGGATCGCAAGACCATCGGAAGAAATTTGCAAGGATGATGATGATGATGATGATGATG 2432
Db 501 ARETTTVDSCNAKSMWRGNWRAKMWAAANNDAGAMDHWTYMMGNNTTMMWRRAKWM 442

Qy 2433 TTCAAGAAACAGAGATGCTATTGATATTTGCTTGTGACATTTGTAACACAGAGAAAGATGT 2492
Db 441 NMAWCRRAYCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 384

```


Db 189 WMKXGRRRATSRYGMMSSMYGASGRMSMCSASTRMSSASCMYMMWSAGSYASCWKM 248
Qy 581 CGGTGAGTTACTGAGCAAACTCTCTCCAGCTCTCTTGCGCGCTCCCGGATTGCTT 640
Db 249 SKYRCAKWCSTYSWYWRASWKSRYCAWSRKGSKCCWYSRKSGSKCTCWWGSCCCCGC 308
Qy 641 CAGAAAGTGTCTAAATTTGGAGCC 662
Db 309 CAGCAGCAGCAGGTGCCAGC 330

RESULT 9

US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 1.4%; Score 41.4; DB 4; Length 399;
Best Local Similarity 19.5%; Pred. No. 0.077;
Matches 56; Conservative 114; Mismatches 110; Indels 7; Gaps 1;

Qy 2401 TGAAGGATGGATGTCAGGTTCCAGAAACGAATTCAGAAACAGAGATGCTATGAGATA 2460
Db 391 TGAAGGTTTGGAGCTATGAGCATCAATCAGCCAAAGKXWWSGAMKWRGTGGWYMYM 332
Qy 2461 TTGCTTGACATTTGAACACAGAGAAAGATGTTTAATAGACTCCAGGGAACATGCTCAA 2520
Db 331 TSGSYRRTKTSAMMGRAKMKRKKTKMYWMKGGKGGSTYAMRSRGTGRWSYR 272
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTGGCAGCTGTAGTCTAGCTAGCATACAC 2580
Db 271 AMRGSXSGGSYRYRAGYRSSRWSYSAWRRKKMTCKWGRSSGSRSTGYAWYK 212
Qy 2581 TACCTCTTACCTGAGAGGTGTCTTTTAAACAAATCTTGGCAGCTGTCTTTGACATTT 2640
Db 211 KSWCTSRKMYKKRRKKRRKCTST-----KRTCYRSGSTYKWKAYYTKRRKRWTR 159
Qy 2641 TTTTGTAGAGGAAATGTAACCTTGATCTAGTTAATTTTCTTTT 2687
Db 158 TTYYYKSYMSMKTKRMTAYYTKRWKTRTKWTCTCMCKCTY 112

RESULT 10

US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141

; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEI promoters
US-09-806-708B-22

Query Match 1.4%; Score 40.8; DB 4; Length 1141;
Best Local Similarity 9.0%; Pred. No. 0.26;
Matches 65; Conservative 291; Mismatches 363; Indels 4; Gaps 1;

Qy 2175 CCCCATCTGATCTCATTTGTGGCCATTTCTGAGTGTCAACCTTCTTCTTGAAGTGTCT 2234
Db 53 MSXSRKWTWARMYCKYRRWYNNKSRWKGWYKKYKKBANNTSBRVHARRKMDKMTAYBM 112
Qy 2235 AGCTGTGGGAGCCCAAGAAAGTGAATGCCAGACTCGGTACAGCGTGACGTTCTTCTG 2294
Db 113 TMTNKGKGTGWRHRYWRWRAMBTVDDHYVYVANNNAWTTWCMMDKDKRTRWKKNNNA 172
Qy 2295 GAGGAAGACACACACATGTCAAAAGAAAGAAAGTAACTTCTTCTTGAAGTGTCTAGT 2354
Db 173 TGWDDDTKYHMMNNNGCBIVTWVRYKTDWDWSBKRMNYGMBWKNWSYDVYVWVWDD 232
Qy 2355 ATTATATAAGTGTGATGAGATCGCAAGGACCATCGGGAAGAAATTTGCAAGGATGGAT 2414
Db 233 MCKRKVRVRVTRGRMRNTVMWBTARRRRYNNNGTBTAMAYRRTWMTNNNNNNKAMCKRA 292
Qy 2415 GCAGGTTCCAGAAACGAATTCAGAAACAGAAAGATGCTATTGATGATTGCTGACATTTG 2474
Db 293 KYGMWRABVNSTCTTWKSKTTRTSCWANNCRAGDANKDHWKWKWSAAMGVYNNNNN 352
Qy 2475 AACACAGAGAAAGATGTTTAATAGACTCCAGGGAACATGCTTCAAGGACTTAATATTC 2534
Db 353 NNWYKKARHBAWDVWHSWKWHAHAHYSRKWTBYKRTVNNNNNGTTWKKEMWA 412
Qy 2535 TCAAGCAATGTTGAGTTGGCAGCTGTAGTCTAGCTAGCATACACTCTTCTTACCTGA 2594
Db 413 WYKMDMDWBGTYNNNNNGRTYVGTGNKMMYTYKWKANNCKWRAWDHKTCTHNNTTW 472
Qy 2595 GAGGTGCTTTTAAA-----AACAAATCTTGGCAGCTGTCTTGGACATTTTCTTTT 2650
Db 473 WKKTYYNNCYWKSMTNGSKSHRBAAYVYTWMMWRYAHANNNNNDYMWKACTWYKYBV 532
Qy 2651 AGAAATTAACCTTGGATCTAGTTAATTTTCTTCTTCAACATATCCCACTCAGAAAC 2710
Db 533 CSKWNWYAAWYTKSSWNYTSRYRYKTNNSWRSWSDTRSMGRANNYARABHYGYKWNTR 592
Qy 2711 ATTCAGGTTTGAAGCCGCTGATTAATGAAGATGAACCTAGTGTGATTTCTTAATCTCC 2770
Db 593 WBSWHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGNNNNNNNNNNNN 652
Qy 2771 CTTTTTGTAGTTAGTGTGCTTTTAAATGCTCTTTGCTGTCATGAGTGAAGGG 2830
Db 653 NNATCARDYYAASRYAMANAKEYYKGAANNAYYTHANNWGWGWNNAVDTTRTWKNN 712
Qy 2831 GACCTTTTGTAGTTGTCAITTTTGCACCTTCAAACTTATTTTCTTGAAACAATATTTA 2890
Db 713 NNNNAGTWKNNNNNNNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAK 772
Qy 2891 TAG 2893
Db 773 NNG 775

RESULT 11

US-09-949-016-15525
; Sequence 15525, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15525
; LENGTH: 117807
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-15525

Query Match          1.4%; Score 40.4; DB 4; Length 117807;
Best Local Similarity 51.1%; Pred. No. 11;
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY      2755 TGATTTCTAATCCTCCCTTTTTTGGATTAGTTGGAATGCCTTTAAAGTCCTTTGGCTG 2814
DB      46405 TGGTATCTCAATGTGTGTGTTTTTTTTTTGGTTTGCTTTTTTAGTTTATTAGTA 46464

QY      2815 CATGAGGTGMAAGGGACCTTTTTTGAGTTGTCATTTTGCACCTTCCAACAATTTTCT 2874
DB      46465 TTTTGTCCCCCATGCATCAATGTGTTTGAATTTGCATTTTCTTAAAACCTTGGA 46524

QY      2875 TGGAAACAATAFTTATAGGCTTAAAGCCCATTTCATTTCTAATCTAAATTAATGTGT 2934
DB      46525 TGGTGAGCATCTTTTCATGTGCTCTTTGGCCATTTGTATGTCTCTTTCATGTTTGGTTG 46584

QY      2935 CCTATC 2940
DB      46585 TATAGC 46590


RESULT 12
US-09-949-016-3573
; Sequence 3573, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3573
; LENGTH: 1468
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-3573

Query Match          1.4%; Score 40.2; DB 4; Length 1468;
Best Local Similarity 55.3%; Pred. No. 0.47;
Matches 78; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY      1688 TATATTTGGAGGTGCATCCAGTCACCTGGAAACAAGTTCTGATCCAGAAGGTGAGGAT 1747
DB      1004 TATTTTACTGGAGAAGCTATTGAAGATGATGATGATTTATGATGAAGAAGGTGAAGAA 1063

QY      1748 TGGGATGAGGAAGCTGAGGATGATGGTTTTTGTAGTGTAGTCACTCTCAGACTCAGAC 1807
```



```

RESULT 14
US-09-919-039-334
; Sequence 334, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kasear, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 334
; LENGTH: 4252
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 995575.17
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1730, 1747, 1751, 1763, 1769, 1771-1772, 1778
; OTHER INFORMATION: a, t, c, g, or other
US-09-919-039-334

```

[illegible]

RESULT 15
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:

```

; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZspt-Fls
; US-08-232-463-14

      Query Match          1.13%; Score 39.6; DB 1; Length 7218;
      Best Local Similarity 6.43; Pred. No. 2.3;
      Matches 27; Conservative 208; Mismatches 187; Indels 0; Gaps 0;

Qy 414 CGGGGACAGCGCGGATCGCGAAACGGGCTTGGCCCTCGGGCGGGCTTCGGTTCCTGGCCAC 473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1016 CGCTCACAAGATTAAATTCGAGCTTGGCTGCAGGTCGAGGGAGCTTCGATATTTTTTYY 1075

Qy 474 CTTTTTCCCTCGGGGATCGCAAGCAGGCTCTTCTAAGTTCGACGCCTCTTGGCCCGG 533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135

Qy 534 AAAACTCCGGGAACCCACACTGTTCTCTCTGCCAGCCGAGACTCGGGTCAGTTTACT 593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1136 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1195

Qy 594 GAGACGAATCTGCTCTCCAGCTCCTTGGCGCGCTCCCGGATGCTTCAGAGGTGCTAA 653
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1196 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1255

Qy 654 TTGGGACCAACTTTTCGGTGGAAATGTTTCCGACCAGATGGCTAGATTTTCTGGAGTCT 713
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1256 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1315

Qy 714 ACAGCGCCCTGAGAGCCCTGAAGGGACGGGAGAAACACGCGCCCCACAGCGCAGAAAT 773
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1316 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375

Qy 774 CTTTGAGTTCGTCGAGCTCGACTCTCAGACCCCTCGGTCCAGCTCCCTTGATGGC 833
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1376 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1435

Qy 834 TA 835
      |||
Db 1436 TA 1437

Search completed: September 16, 2005, 09:16:56
Job time : 445.026 secs

```

Search completed: September 16, 2005, 09:16:56
Job time : 445.026 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query #		Length	DB	ID	Description
		Match	Length				
1	2942	100.0	2942	19	US-10-650-482-1		Sequence 1, Appl.
2	2185.8	74.3	2324	9	US-09-925-299-168		Sequence 168, App
3	2185.8	74.3	2324	10	US-09-925-299-168		Sequence 168, App
4	1324	45.0	5468	19	US-10-650-482-3		Sequence 3, Appl
5	828.6	28.2	883	18	US-10-296-115-224		Sequence 224, Appl
6	435.8	14.8	465	22	US-10-755-149-3997		Sequence 3997, Ap
7	265.8	9.0	558	9	US-09-878-178-2040		Sequence 2040, Ap

```

81 40.2 1.4 2410 10 US-09-814-353-21256 Sequence 21256, A
82 40.2 1.4 2908 19 US-10-734-564-22 Sequence 22, App1
83 40.2 1.4 2908 21 US-10-887-553A-345 Sequence 345, App
84 40.2 1.4 3549 21 US-10-956-157-4786 Sequence 4786, Ap
85 40.2 1.4 3585 21 US-10-956-157-1264 Sequence 1264, Ap
86 40.2 1.4 4252 10 US-09-919-039-334 Sequence 334, App
87 40.2 1.4 6048 18 US-10-221-714A-336 Sequence 336, App
88 40 1.4 752 17 US-10-264-049-679 Sequence 679, App
89 40 1.4 6816 20 US-10-723-860-7676 Sequence 7676, Ap
90 40 1.4 7341 17 US-10-221-613-357 Sequence 357, App
91 40 1.4 18624 15 US-10-311-455-1676 Sequence 1676, App
92 40 1.4 61020 18 US-10-221-714A-513 Sequence 513, App
93 39.8 1.4 6167 17 US-10-221-613-292 Sequence 292, App
94 39.8 1.4 6169 18 US-10-221-714A-91 Sequence 91, App1
95 39.8 1.4 16766 15 US-10-311-455-2130 Sequence 2130, Ap
96 39.6 1.3 20486 15 US-10-240-485-163 Sequence 163, App
97 39.4 1.3 453 21 US-10-706-635-56 Sequence 56, App1
98 39.4 1.3 1337 13 US-10-027-632-123572 Sequence 123572,
99 39.4 1.3 1337 13 US-10-027-632-123573 Sequence 123573,
100 39.4 1.3 1337 17 US-10-027-632-123572 Sequence 123572,

c
; Sequence 1, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/1L712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2942
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-650-482-1

Query Match 100.0%; Score 2942; DB 19; Length 2942;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2942; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTTGGGCTTCGCTTCCACCGCACCAGCCGGCCCTACCCAGTCTCTTCGGGTATCGCGTTG 60
Db 1 ATTTTGGGCTTCGCTTCCACCGCACCAGCCGGCCCTACCCAGTCTCTTCGGGTATCGCGTTG 60

Qy 61 CTCAGGGGCTTTTCAACCCCTCTGTCTCAGTCGGAACCAATCCCGAGGCGCTGGGGGACT 120
Db 61 CTCAGGGGCTTTTCAACCCCTCTGTCTCAGTCGGAACCAATCCCGAGGCGCTGGGGGACT 120

Qy 121 CCTATCCATGGTGTGAAGCGTCGAGCCGATAGGAAACCTTCCTTCCCGCCAGGATGGA 180
Db 121 CCTATCCATGGTGTGAAGCGTCGAGCCGATAGGAAACCTTCCTTCCCGCCAGGATGGA 180

Qy 181 AGTCGCATCAGTCGCGGCTTATTCGGGGGCTGTCTTCCTGTGTCTGTCGCGCCGCTG 240
Db 181 AGTCGCATCAGTCGCGGCTTATTCGGGGGCTGTCTTCCTGTGTCTGTCGCGCCGCTG 240

Qy 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCTCGAGCAGGAC 300
Db 241 CCGCATTCGCTGCCCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCTCGAGCAGGAC 300

Qy 301 GCCACCGCTGGGCAAGGCCGAGACTCTGTAGGCTTCTTCGAAATCCCGTGCACCTCAGC 360

```

ALIGNMENTS

```

Db 301 GCCACCGCTGGGCAAGGCCGAGACTCTGTAGGCTTCTTCGAAATCCCGTGCACCTCCAGC 360
Qy 361 CGCTGAGCGCGCGGCGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCCGGGAC 420
Db 361 CGCTGAGCGCGCGGCGCTTACCTGAGAGACTGTCAAGAAAAGAGATGGAGCCGGGAC 420
Qy 421 AGGCGGATCGCGGAAAACGGCTTGGCCCTCGGGCGGGCTTCGGGTTCTGGCCACCCCTTTT 480
Db 421 AGGCGGATCGCGGAAAACGGCTTGGCCCTCGGGCGGGCTTCGGGTTCTGGCCACCCCTTTT 480
Qy 481 CCCTCGGCGATCGCAAGCAGGCTCTTAAAGTTCCGACGCGCTTGGCCCGGAAAACCTC 540
Db 481 CCCTCGGCGATCGCAAGCAGGCTCTTAAAGTTCCGACGCGCTTGGCCCGGAAAACCTC 540
Qy 541 CGGGAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTACTTCGACGAA 600
Db 541 CGGGAACCCACACTGCTTCTCTGCGCCAGCCGAGACTCGGGTCAGTTACTTCGACGAA 600
Qy 601 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCGTTTCAAGAGTGCTTAATTTGGAG 660
Db 601 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCGTTTCAAGAGTGCTTAATTTGGAG 660
Qy 661 CCAACTTTTCGGTGGAAATGTTTCGACAGATGGCTAGATTTTGTCTGGAGTCTACAGCGC 720
Db 661 CCAACTTTTCGGTGGAAATGTTTCGACAGATGGCTAGATTTTGTCTGGAGTCTACAGCGC 720
Qy 721 CCTGAGAGCCCTGAAGGAGCGGAGAAACAGCCGCCCCACAGCGCAGAGAAATCTTTGAG 780
Db 721 CCTGAGAGCCCTGAAGGAGCGGAGAAACAGCCGCCCCACAGCGCAGAGAAATCTTTGAG 780
Qy 781 TTGCTGCTCAGCTCGACTCTCTCAGACCCCTCGGTCCAGTCCCTTGAITTTGGCTAGAGGA 840
Db 781 TTGCTGCTCAGCTCGACTCTCTCAGACCCCTCGGTCCAGTCCCTTGAITTTGGCTAGAGGA 840
Qy 841 GGGGATCCACTGGCAATACTCGCCGCCAGACTTAAATTTGGAGCTTAAAGCCAAAGGAAG 900
Db 841 GGGGATCCACTGGCAATACTCGCCGCCAGACTTAAATTTGGAGCTTAAAGCCAAAGGAAG 900
Qy 901 TGCTTTGGAGCCCTGACGACAGAGCTTTTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960
Db 901 TGCTTTGGAGCCCTGACGACAGAGCTTTTCTTAGAGCAGCAGCTGTGGGAGTGGAGCT 960
Qy 961 GTTGCCCGAGTAGCCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC 1020
Db 961 GTTGCCCGAGTAGCCTTCAATCCCGTCTGTACTCTTAACCGGGAACTTGGCTCTTCGCCCTC 1020
Qy 1021 TGGGCTCTTAAACATTCGATAGCAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 1080
Db 1021 TGGGCTCTTAAACATTCGATAGCAATTTTCAGTGTGGTATCTTATTTGCTGAACCC 1080
Qy 1081 TTCTTACCTCGACTGCTTTTCTTAGGCTAGAGTCAAGTATCAGAACAGTGTGAAATAG 1140
Db 1081 TTCTTACCTCGACTGCTTTTCTTAGGCTAGAGTCAAGTATCAGAACAGTGTGAAATAG 1140
Qy 1141 CGAGGTAGTGGCTTCCAGACACTAACCCAGAGAGCAGCTGCTGAGAGAGGACCATTG 1200
Db 1141 CGAGGTAGTGGCTTCCAGACACTAACCCAGAGAGCAGCTGCTGAGAGAGGACCATTG 1200
Qy 1201 TCATTCGCCAGCGCTGAGTGCAGAACTCATTCGGCCCTCGTGGCAGGGATGTCCACCTCT 1260
Db 1201 TCATTCGCCAGCGCTGAGTGCAGAACTCATTCGGCCCTCGTGGCAGGGATGTCCACCTCT 1260
Qy 1261 TTCTTACGGAAGGCTACCCAGAAATTCACCATCTTCGATCAAAACGGCTGGAATTCCTTCA 1320
Db 1261 TTCTTACGGAAGGCTACCCAGAAATTCACCATCTTCGATCAAAACGGCTGGAATTCCTTCA 1320
Qy 1321 ACAGGCTTAAACAGGGGCAAGATTTTACCCCTCGACGAGTAATGGCTTACACAGCCT 1380
Db 1321 ACAGGCTTAAACAGGGGCAAGATTTTACCCCTCGACGAGTAATGGCTTACACAGCCT 1380
Qy 1381 GAGGAGGAGACACAGCCTTCTCGGATGGATCCAAACACTGCAGAGATAACCCACACA 1440
Db 1381 GAGGAGGAGACACAGCCTTCTCGGATGGATCCAAACACTGCAGAGATAACCCACACA 1440

```

```
Qy 1441 GTTTGTTCTCTGCTGAGACATTTCTCTGGAAACACCCAGGAATCCACTGAAGAAAAAT 1500
Db 1441 GTTTGTTCTCTGCTGAGACATTTCTCTGGAAACACCCAGGAATCCACTGAAGAAAAAT 1500
Qy 1501 AGAATTATTAACACAGAGGTTCCACTGCTTTTGGAGAAGAGAGCCCTCTGAGGCGTG 1560
Db 1501 AGAATTATTAACACAGAGGTTCCACTGCTTTTGGAGAAGAGAGCCCTCTGAGGCGTG 1560
Qy 1561 TCCATCTAGTGAGATACCTATGGAAAAGAGCCCTGGAGAGGCGCGAATAAGTGTAGTTGA 1620
Db 1561 TCCATCTAGTGAGATACCTATGGAAAAGAGCCCTGGAGAGGCGCGAATAAGTGTAGTTGA 1620
Qy 1621 TTACTCATACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTTTGTAGTACAACACT 1680
Db 1621 TTACTCATACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCCAGCTTTGTAGTACAACACT 1680
Qy 1681 GATAGATTATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTTCTGATCAGAAGG 1740
Db 1681 GATAGATTATATTTTGGAGGTGATCCAGTGACCTGGAAACAAGTTCTGATCAGAAGG 1740
Qy 1741 TGAGGATGGGATCAGGAAGCTGAGGATGATGGTTTTTGTATGATGATGATGATGATGATG 1800
Db 1741 TGAGGATGGGATCAGGAAGCTGAGGATGATGGTTTTTGTATGATGATGATGATGATGATG 1800
Qy 1801 CTCAGACCTTGAACAAGACCCCTGAAGGCTTTCACCTTTTGGAACTCTTTTCTGCAGTGTAGA 1860
Db 1801 CTCAGACCTTGAACAAGACCCCTGAAGGCTTTCACCTTTTGGAACTCTTTTCTGCAGTGTAGA 1860
Qy 1861 TCCTTATTAATCCCAAGACTTTACAGCAACAAATTCAGAGCTGCTGCCAGAAATGTTCTCTGA 1920
Db 1861 TCCTTATTAATCCCAAGACTTTACAGCAACAAATTCAGAGCTGCTGCCAGAAATGTTCTCTGA 1920
Qy 1921 AGAGCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATTCCTCCCA 1980
Db 1921 AGAGCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATTCCTCCCA 1980
Qy 1981 GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTTGGGAGGAAGATGATCGGGAATC 2040
Db 1981 GTCTGGAAGCCTTCTCTGAGACCCCTGAGCATAGTTCTTGGGAGGAAGATGATCGGGAATC 2040
Qy 2041 TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATCTGTAATTCGTGATGACCC 2100
Db 2041 TAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGAATCATCTGTAATTCGTGATGACCC 2100
Qy 2101 CTACAACCCCTTTAAATTTTAAAGGCTCTCTTTTCAAACATCAGGGGAAAATGAGAAGGCTG 2160
Db 2101 CTACAACCCCTTTAAATTTTAAAGGCTCTCTTTTCAAACATCAGGGGAAAATGAGAAGGCTG 2160
Qy 2161 TCGTGACTCAAAGACCCCATCTGAGTCCATTTGTGGCCATTTTCTGAGTGTCAACCTTTACT 2220
Db 2161 TCGTGACTCAAAGACCCCATCTGAGTCCATTTGTGGCCATTTTCTGAGTGTCAACCTTTACT 2220
Qy 2221 TTCTTGTAAAGTGAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Db 2221 TTCTTGTAAAGTGAGCTGTTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCG 2280
Qy 2281 TGAGCTTCTTCTGAGGAGACACACACATGTCAAAGAAAAGGTAACTTTCCTTTGA 2340
Db 2281 TGAGCTTCTTCTGAGGAGACACACACATGTCAAAGAAAAGGTAACTTTCCTTTGA 2340
Qy 2341 AGAAGTTACTGAGTATTATATAAGTGTGTGATGAGGATCGCAAGGACCATGGGAAGATT 2400
Db 2341 AGAAGTTACTGAGTATTATATAAGTGTGTGATGAGGATCGCAAGGACCATGGGAAGATT 2400
Qy 2401 TCGAAGGATGATGACAGGTTCCAGAAACGAATTCAGAAACAGAAATGCTATTGGATA 2460
Db 2401 TCGAAGGATGATGACAGGTTCCAGAAACGAATTCAGAAACAGAAATGCTATTGGATA 2460
Qy 2461 TTGCTTGACATTTGAAACACAGAGAAGAAATGTTTTAATAGATCCAGGGGAACATGCTTCAA 2520
Db 2461 TTGCTTGACATTTGAAACACAGAGAAGAAATGTTTTAATAGATCCAGGGGAACATGCTTCAA 2520
```

```
Qy 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Db 2521 AGGACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCCTGTAGTCTCTAGCTAGCATACAC 2580
Qy 2581 TACCTCTTACCTCAGAGGTGCTCTTTTAAAAAACAAATCTTGGCAGCTGTCTTTGACATTT 2640
Db 2581 TACCTCTTACCTCAGAGGTGCTCTTTTAAAAAACAAATCTTGGCAGCTGTCTTTGACATTT 2640
Qy 2641 TTTTCTTTAGAGGAAATGTAACCTTGGATCTAGTTAAATTTTTTTTTTGTGCAACATATCCC 2700
Db 2641 TTTTCTTTAGAGGAAATGTAACCTTGGATCTAGTTAAATTTTTTTTTTGTGCAACATATCCC 2700
Qy 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAGGATGAATAGTGTGATTT 2760
Db 2701 ACTCAGAAACATTCAGGTTTGAAGCCAGCCCTGATTAATGAGGATGAATAGTGTGATTT 2760
Qy 2761 CTAACTCTCCCTTTTGTATTTAGTTGGATGTCCTTTTAAATCTCTTTCSCCTGCAATGAG 2820
Db 2761 CTAACTCTCCCTTTTGTATTTAGTTGGATGTCCTTTTAAATCTCTTTCSCCTGCAATGAG 2820
Qy 2821 GTGGAAGGGGACCTTTTGTAGTTGTCAATTTTGCACCTTTCAAACCTTATTTTCTTGGAAA 2880
Db 2821 GTGGAAGGGGACCTTTTGTAGTTGTCAATTTTGCACCTTTCAAACCTTATTTTCTTGGAAA 2880
Qy 2881 ACAATATTTATAGGGCTTAAAGCCCATTTTCAATTTCTAAATTAATGTTGTGCTATC 2940
Db 2881 ACAATATTTATAGGGCTTAAAGCCCATTTTCAATTTCTAAATTAATGTTGTGCTATC 2940
Qy 2941 TG 2942
Db 2941 TG 2942
```

RESULT 2

```
US-09-925-299-168
; Sequence 168, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-168
```

```
Query Match 74.3%; Score 2185.8; DB 9; Length 2324;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 2249; Conservative 6; Mismatches 36; Indels 6; Gaps 4;
```

```
Qy 456 GCTTCGGTCTCGGCCACCCCTTTTCCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCC 515
Db 32 GCTCAGGGGCTTTTCAACCCCTCTGTGAGTGGCGATCGCAAGCAGGCTCTTCTAAGTTCC 91
Qy 516 CGACGCTCTTGTGGCCCGGAAAACCTCCGGGAAACCCCACTGCTTTTCTCTGCCCCAGCCCCG 575
Db 92 CGACGCTCTTGTGGCCCGGAAAACCTCCGGGAAACCCCACTGCTTTTCTCTGCCCCAGCCCCG 151
Qy 576 AGACTCGGGTCAGTTACTGGAAGAAATGCTCTCCAGCTCTTTGCGCGCGCTCCCCGGAT 635
Db 152 AGACTCGGGTCAGTTACTGGAAGAAATGCTCTCCAGCTCTTTGCGCGCGCTCCCCGGAT 211
Qy 636 TGCTTCAGAAAGGTGCTAAATTTGAGGCCAACTTTTCGGTGGAAATGTTTCCGACCATGGC 695
```

Db 212 TGCTTCAGAAAGTGCTAAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACCAAGATGCG 271
 Qy 696 TAGATTTTCTCGAGTCTACAGGCCCTGAGAGCCCTGAAGGACGGGAGAAACACGCG 755
 Db 272 TAGATTTTCTCGAGTCTACAGGCCCTGAGAGCCCTGAAGGACGGGAGAAACACGCG 331
 Qy 756 CCCCCACAGCGCAGAAATCTTTGAGTTTCGCTCGAGCTCGACTCCCTCAGACCCCTCGGTCA 815
 Db 332 CCCCCACAGCGCAGAAATCTTTGAGTTTCGCTCGAGCTCGACTCCCTCAGACCCCTCGGTCA 391
 Qy 816 CAGTCCCTTCGATTTGGCTAGAGGAGGGATCCATGCGCAATACATGCGCCCCAGACCTTAA 875
 Db 392 CCAAGTCCCTTCGATTTGGCTAGAGGAGGGATCCATGCGCAATACATGCGCCCCAGACCTTAA 451
 Qy 876 AATTGGAGCTTAAGGCCAAGGAGGAGTCTTTGGAGCCCTGCAGCAGAGGCTTTTCCTTAG 935
 Db 452 AATTGGAGCTTAAGGCCAAGGAGGAGTCTTTGGAGCCCTGCAGCAGAGGCTTTTCCTTAG 511
 Qy 936 AGCAGCAGCTGTGGGAGGTGGAGCTGTTCGCCAGTAGCCTTCAATCCCGTCTGTACTCTA 995
 Db 512 AGCAGCAGCTGTGGGAGGTGGAGCTGTTCGCCAGTAGCCTTCAATCCCGTCTGTACTCTA 571
 Qy 996 ACCGGGAATCTGGCTCTTCGCGCTCTGGGCTCTTAAACATTCOAACGATAGACAAATTTCA 1055
 Db 572 ACCGGGAATCTGGCTCTTCGCGCTCTGGGCTCTTAAACATTCOAACGATAGACAAATTTCA 631
 Qy 1056 GTGTGGTATCCTATTTGCTGAACCTTCTACCTGCACTGCTTTCTTAGGCTAGAGTCA 1115
 Db 632 GTGTGGTATCCTATTTGCTGAACCTTCTACCTGCACTGCTTTCTTAGGCTAGAGTCA 691
 Qy 1116 GCTATCAGAACAGTGAATAGCGAGGTAGTGGCTTCAGACACTAACCCAGAGA 1175
 Db 692 GCTATCAGAACAGTGAATAGCGAGGTAGTGGCTTCAGACACTAACCCAGAGA 751
 Qy 1176 GCAGCTGCCTGAGAGAGACCAATGTTCATCCAGCGCTGAGTGCAGAACTCATTCGG 1235
 Db 752 GCAGCTGCCTGAGAGAGACCAATGTTCATCCAGCGGTGARTGCAGAACTCATTCGG 811
 Qy 1236 CTTCTGGCAGGAGTGTCCACCTCTTCTACGGAAGGCTACCAAGAAATTCACCATCTTC 1295
 Db 812 SCTCTGGCAGGAGTGTCCACCTCTTCTACGGAAGGCTACCAAGAAATTCACCATCTTC 871
 Qy 1296 GCATGAACCGCTGGAATCTTCAACAGGCTTAAACAGGGCAAGATTTACCCACCCCTG 1355
 Db 872 GCATGAACCGCTGGAATCTTCAACAGGCTTAAACAGGGCAAGATTTACCCACCCCTG 931
 Qy 1356 ACCAGATAATGGCTACCAAGCTGAGGAGGAAACAGCCCTCTCCGATGGATCCAA 1415
 Db 932 ACCAGATAATGGCTACCAAGCTGAGGAGGAAACAGCCCTCTCCGATGGATCCAA 991
 Qy 1416 AACCTGCAGAGATAACCCAAACACAGTTTGTCTGCTGCTGGAGACATTCCTGGAAACA 1475
 Db 992 AACCTGCAGAGATAACCCAAACACAGTTTGTCTGCTGCTGGAGACATTCCTGGAAACA 1051
 Qy 1476 CCCAGGAATCCACTGAAGAAAAATAGAAATTAATTAACACAGAGTTCCACTTGTCTGG 1535
 Db 1052 CCCAGGAATCCACTGAAGAAAAATAGAAATTAATTAACACAGAGTTCCACTTGTCTGG 1111
 Qy 1536 AAGAGAGAGCCCTCTGAGGCTGTCCATCTAGTGAGATACCTATGAAAGAGGCTG 1595
 Db 1112 AAGAGAGAGCCCTCTGAGGCTGTCCATCTAGTGAGATACCTATGAAAGAGGCTG 1171
 Qy 1596 GAGAGGCCGAATAAGTGTAGTTGATTAATCTACTAGAGGTGACCTTCCCATTTCTG 1655
 Db 1172 GAGAGGCCGAATAAGTGTAGTTGATTAATCTACTAGAGGTGACCTTCCCATTTCTG 1231
 Qy 1656 CCAGACCAAGCTTGTAGTAAACAACTGATAGATTAATTTTGGAGGATGCATCCAGTGACC 1715
 Db 1232 CCAGACCAAGCTTGTAGTAAACAACTGATAGATTAATTTTGGAGGATGCATCCAGTGACC 1291
 Qy 1716 TGGAAACAAGTTCTGATCCAGAGGATGAGGATTTGGATGAGGAGCTGAGGATGATGTT 1775
 Db 1292 TGGAAACAAGTTCTGATCCAGAGGATGAGGATTTGGATGAGGAGCTGAGGATGATGTT 1351

Qy 1776 TTGATAGTAGTGTCACTGTGCACTCAGACCTTGAAACAGACCCCTGAAGGGCTTCACC 1835
 Db 1352 TTGATAGTAGTGTCACTGTGCACTCAGACCTTGAAACAGACCCCTGAAGGGCTTCACC 1411
 Qy 1836 TTTGGAATCTTTCTGAGTGTAGATCCTTATAATATCCCAAGAACTTTTACAGCAAAATTC 1895
 Db 1412 TTTGGAATCTTTCTGAGTGTAGATCCTTATAATATCCCAAGAACTTTTACAGCAAAATTC 1471
 Qy 1896 AGACTGTGCGCAGAAATTTGTTCTGGAAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1955
 Db 1472 AGACTGTGCGCAGAAATTTGTTCTGGAAGAGCTTCTGATTCAGAGAGGATTTGTCTGGCA 1531
 Qy 1956 AGTCTGATCTAGAGAAATTTCTCCAGTGTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 2015
 Db 1532 AGTCTGATCTAGAGAAATTTCTCCAGTGTGGAAGCCCTTCTGAGACCCCTGAGCATAGTT 1591
 Qy 2016 CTGGGAGGAAGATGATCTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGA 2075
 Db 1592 CTGGGAGGAAGATGATCTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAAACTGTGGA 1651
 Qy 2076 ACTCATTTCTGTAATCTGATGACCCCTTACACCCCTTTAAATTTTAAAGGCTCCTTTTCAA 2135
 Db 1652 ACTCATTTCTGTAATCTGATGACCCCTTACACCCCTTTAAATTTTAAAGGCTCCTTTTCAA 1711
 Qy 2136 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTGG 2195
 Db 1712 CATCAGGGGAAAAATGAGAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTGG 1771
 Qy 2196 CCATTTCTGAGTGTCAACCTTTACTTTCTGTAAGGTGACGCTGTTGGGAGGCAAGAA 2255
 Db 1772 CCATTTCTGAGTGTCAACCTTTACTTTCTGTAAGGTGACGCTGTTGGGAGGCAAGAA 1831
 Qy 2256 GTGAATCTCCAGACTCGGTACAGCTGACGTTCTTTCTGGAGGAAGACACACATGTCA 2315
 Db 1832 GTGAATCTCCAGACTCGGTACAGCTGACGTTCTTTCTGGAGGAAGACACACATGTCA 1891
 Qy 2316 AAAGAAAAAGGTAACTCTCTTGAAGAAGTTACTGAGTATTATATAAGTGGTGATGAGG 2375
 Db 1892 AAAGAAAAAGGTAACTCTCTTGAAGAAGTTACTGAGTATTATATAAGTGGTGATGAGG 1951
 Qy 2376 ATCCCAAGGACCATGGAAGAAATTTGCAAGGGATGGAATGCAAGGTTCCAGAAACGAATTC 2435
 Db 1952 ATCCCAAGGACCATGGAAGAAATTTGCAAGGGATGGAATGCAAGGTTCCAGAAACGAATTC 2011
 Qy 2436 AAGAAACAGAGAGTGTATTGGATATTGCTTGACATTTGAAACACAGAGAAAGAAATTTTA 2495
 Db 2012 AAGAAACAGAGAGTGTATTGGATATTGCTTGACATTTGAAACACAGAGAAAGAAATTTTA 2071
 Qy 2496 ATAGACTCCAGGGAAACATGCTTCAAGGACTTTAATGTTCTCAAGCAATGTTGAGTTGGCA 2555
 Db 2072 ATAGACTCCAGGGAAACATGCTTCAAGGACTTTAATGTTCTCAAGCAATGTTGAGTTGGCA 2131
 Qy 2556 GCCTGTAGTCTAGTAGCATACACTACCTTACCTGAGAGGTTCTTTTAAAAACAA 2615
 Db 2132 GCCTGTAGTCTAGTAGCATACACTACCTTACCTGAGAGGTTCTTTTAAAAACAA 2191
 Qy 2616 TCTTGGCAGCTGCTCTTTGACATTTTTTTTTTTTATAGAGGAAATGTAACCTTGGATCTAGTTT 2675
 Db 2192 TCTTGGCAGCTGCTCTTTGACATTTTTTTTTTTTATAGAGGAAATGTAACCTTGGATCTAGTTT 2247
 Qy 2676 AATTTTTTTTTTGTCAACATATCCCACTCCAGAAACATT-CAGGTTTGAAGCCAGCCCT-G 2733
 Db 2248 TTTTTTTTTTTTGTCAACATATCCCACTCCAGAAACATTCCAGGTTTGAAGCCAGCCCTGG 2307
 Qy 2734 ATAATGAAGGATGAAT 2750
 Db 2308 ATAATGAAGGATGAAT 2324

Db	752	GCAGCTGCCTGAGAGAGGACCAATTGTCATCCCCAGCCGGYTGARTGCAGAACTCATTTCCGG	811
Qy	1236	CTTCGTGGCAGGATGTCACCTCTTTCTTACGGAAGGCCTACAGAAATTTACCATCTTC	1295
Db	812	CTTCGTGGCAGGATGTCACCTCTTTCTTACGGAAGGCCTACAGAAATTTACCATCTTC	871
Qy	1296	GCATGAACCGCTGGAAATTCCTTCAACAGACTAACAGGGGCAAGATTTTACCCACCCCTG	1355
Db	872	GCATGAACCGCTGGAAATTCCTTCAACAGGCTAGCAAGGGGCAAGATTTTACCCACCCCTG	931
Qy	1356	ACCAGGATAATGGCTACCAACAGCCTCGAGGAGGAACAACAGCCTTCTCCGGATGGATCCAA	1415
Db	932	ACCAGGATAATGGCTACCAACAGCCTCGAGGAGGAACAACAGCCTTCTCCGGATGGATCCAA	991
Qy	1416	AACACTGCAGAGATAACCCAAACAGTTTGTCTGCTGTGAGAGACATTTCTCTGGAAACA	1475
Db	992	AACACTGCAGAGATAACCCAAACAGTTTGTCTGCTGTGAGAGACATTTCTCTGGAAACA	1051
Qy	1476	CCCAGGAATCCACTGAAGAAATAAGAAATTAATTAACACAGAGGTTCCACTTCGTTGG	1535
Db	1052	CCCAGGAATCCACTGAAGAAATAAGAAATTAATTAACACAGAGGTTCCACTTCGTTGG	1111
Qy	1536	AAGAAAGAGAGCCCTTCTGAGGGCTGTCATCTAGTGAGATACCTATATGGAAAAAGGAGCCTG	1595
Db	1112	AAGAAAGAGAGCCCTTCTGAGGGCTGTCATCTAGTGAGATACCTATATGGAAAAAGGAGCCTG	1171
Qy	1596	GAGAGGCCGAATAAGTGATGATTAATCTATACCTAGAGGTGACCTTCCCATTTCTG	1655
Db	1172	GAGAGGCCGAATAAGTGATGATTAATCTATACCTAGAGGTGACCTTCCCATTTCTG	1231
Qy	1656	CCAGACAGCCTGTAGTAAACAAACTGATAGATTAATATTTTGGAGGTGCATCCAGTGACC	1715
Db	1232	CCAGACAGCCTGTAGTAAACAAACTGATAGATTAATATTTTGGAGGTGCATCCAGTGACC	1291
Qy	1716	TGGAAAACAAGTTCTGATCCAGAAAGTTCAGAGTGGGATGAGGAAGCTGAGGATCATGTT	1775
Db	1292	TGGAAAACAAGTTCTGATCCAGAAAGTTCAGAGTGGGATGAGGAAGCTGAGGATCATGTT	1351
Qy	1776	TTGATAGTGATGATCTCACTGTCAGACTCAGACCTTGAAACAGACCCCTGAAAGGGCTTCACC	1835
Db	1352	TTGATAGTGATGATCTCACTGTCAGACTCAGACCTTGAAACAGACCCCTGAAAGGGCTTCACC	1411
Qy	1836	TTTGGAACTCTTTCTGCAGTGTAGATCCTTAATATCCCAAGAACTTTACAGCAACAATTC	1895
Db	1412	TTTGGAACTCTTTCTGCAGTGTAGATCCTTAATATCCCAAGAACTTTACAGCAACAATTC	1471
Qy	1896	AGACTGCTGCCAGAAATGTTTCTTGAAGAGCCTCTGATTCAGAGAAAGATTTGTCTGGCA	1955
Db	1472	AGACTGCTGCCAGAAATGTTTCTTGAAGAGCCTCTGATTCAGAGAAAGATTTGTCTGGCA	1531
Qy	1956	AGTCTGATCTAGAGAAATTCCTCCAGTCTCGAAGCCTTCTCTGAGACCCCTGAGCATAGTT	2015
Db	1532	AGTCTGATCTAGAGAAATTCCTCCAGTCTCGAAGCCTTCTCTGAGACCCCTGAGCATAGTT	1591
Qy	2016	CTGGGAGGAGATGACTGGGAATCTAGTGAGATGAAGCAGAGAGTCTCAAACTGTGGA	2075
Db	1592	CTGGGAGGAGATGACTGGGAATCTAGTGAGATGAAGCAGAGAGTCTCAAACTGTGGA	1651
Qy	2076	ACTCATTTCTGTAAATCTGTATGACCCCTCAACCCCTTAATATTTTAAAGGCTCCTTTTCAAA	2135
Db	1652	ACTCATTTCTGTAAATCTGTATGACCCCTCAACCCCTTAATATTTTAAAGGCTCCTTTTCAAA	1711
Qy	2136	CATCAGGGGAAATAGAAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTCTGG	2195
Db	1712	CATCAGGGGAAATAGAAAAGGCTGTGCTGACTCAAAAGACCCCATCTGAGTCCATTTCTGG	1771
Qy	2196	CCATTTCTGAGTGTACACCTTACTTTTGTATAGGTGCAGCTGTGGGAGGCCAAGAAA	2255
Db	1772	CCATTTCTGAGTGTACACCTTACTTTTGTATAGGTGCAGCTGTGGGAGGCCAAGAAA	1831
Qy	2256	GTGAATGTCAGACTCGGTACAGCGTGACGTTCTTTCTTGAGGAGAGACACACATGTCA	2315
Db	1832	GTGAATGTCAGACTCGGTACAGCGTGACGTTCTTTCTTGAGGAGAGACACACATGTCA	1891

```

QY 2316 AAAAGAAAAGCGTAACTTCTTGAAGAAGTTACTGAGTATTATATAAGTGGTATGAGG 2375
Db 1892 AAAGAAAAGAGTAACCTTCTTGAAGAAGTTACTGAGTATTATATAAGTGGTATGAGG 1951
QY 2376 ATCGCAAAAGGACCATGGGAAGAAATTTGCAAGGGATGAGTGCAGGTTCCAGAAAACGAATTC 2435
Db 1952 ATCGCAAAAGGACCATGGGAAGAAATTTGCAAGGGATGAGTGCAGGTTCCAGAAAACGAATTC 2011
QY 2436 AAGAAACAGAGATGCTATTGATATTGCTTGAACATTTGAAACACAGAGAAAGATGTTTA 2495
Db 2012 AAGAAACAGAGATGCTATTGATATTGCTTGAACATTTGAAACACAGAGAAAGATGTTTA 2071
QY 2496 ATAGACTCCAGGGAACATGCTTCAAGGACATTAATGTTCTCAAGCAATGTTGAGTTGSCA 2555
Db 2072 ATAGACTCCAGGGAACATGCTTCAAGGACATTAATGTTCTCAAGCAATGTTGAGTTGSCA 2131
QY 2556 GCCTGAGTCTAGTAGCATACACTACTCTTACCTGAGAGGTGCTCTTTTAAAAAACAAA 2615
Db 2132 GCCTGAGTCTAGTAGCATACACTACTCTTACCTGAGAGGTGCTCTTTTAAAAAACAAA 2191
QY 2616 TCTGGGAGCTGCTCTTGAACATTTTCTTTTATAGGAAATGTAACCTTGATCTAGTTT 2675
Db 2192 TCTGGGAGCTGCTCTTGAACATTTTCTTTTATAGGAAATGTAACCTTGATCTAGTTT 2247
QY 2676 AATTTTTTTTTTTGCAACATATCCCACTCAGAAACATTT-CAGGTTTGAAGCCAGCCCT-G 2733
Db 2248 TTTTTTTTTTTTTGCAACATATCCCACTCAGAAACATTTCCAGGTTTGAAGCCAGCCCTGG 2307
QY 2734 ATAATGAAGGATGAAT 2750
Db 2308 ATAATGAAGGATGAAT 2324

RESULT 4
US-10-650-482-3
; Sequence 3, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; FILE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5468
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-650-482-3

Query Match 45.0%; Score 1324; DB 19; Length 5468;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1964; Conservative 0; Mismatches 660; Indels 84; Gaps 16;

QY 96 CCATCGCGAGCCGCTGGGGGACTCTCATCCATGGTGTGAAGCGTCGAGCCGACTAGG 155
Db 151 CGTCCGCGCGCGGAGGGGTCTCTCTATGTTGAGGCGATCTCACAGGCGCTAGG 210
QY 156 GAACCTCTCTCC-CGCCAGGATGAAAGTCGATCAGTCGCGCCCTATTGCGCGGCTGT 214
Db 211 AGCTCTCTCTCCAGCGGATGACCTAACCGCGTCCGACCGCTTGC CGGGCCTC 270
QY 215 TCTTCCCTGTGTCTCGCGCCGCTCGCATTTCCCTCTCTGCTGCTTCTGCTGG 274
Db 271 TGGGCGCTCGCGGTGAGCACTGTTGCGGAGCCGCGCTCTCTGGGCTCTCTGCGG 330
QY 275 CTCGAAGATCGGCTTGGAGCAGCGCAACCGCTGGGCAAGCGCGAGACTCTGTAGGCT 334

```

```

Db 331 CGCGGGAATCGGACTGAGTACCCACTCGTGGCTGGCAAGCGGAGAGCTGTGTAGACC 390
QY 335 TCCTCCGAATCCCGTGCACCTTCAGCGCGCTGAGCGCGCGGCTTACTGAGAGACTGTGTC 394
Db 391 TCGGATCCAGCCTCGCTGACGCGCTGAGCTCTGTCTCTCTCTGCTGAGAAAGCGGCC 450
QY 395 AAGAAAAGAGATGAGCGCGGACAGGCGGATCGCGGAAACGCGCTTGGCCCTTCGGCGC 454
Db 451 AAG-GAAGAGATGAGAGACAGGAAACGCAAGGCGCGGAGCGGCTTGGCCCTTCGGCTG 509
QY 455 GGCTTCGCGTTCGGCCACCCCTTTTCCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTC 514
Db 510 GGCTTCCTGTTTCGGCTGCC--TTCTCTTCGGCGATCGCAAGCTGCTCTTCGGAGTTC 566
QY 515 CCNAGCGCTCTTGGCCCGGAAACTCGGGAAACCCACACTGCTTCTTCTTCTGCCCCAGCCC 574
Db 567 CCGCGCGCTTCTCTCGACAAATTCGCGGAA-----CTCGCTCTGCC 611
QY 575 GAGACTCGGTCAGTACTGAGCGAAATCTCTCTCCAGCTCTTTCGCGCGCTCCCGGA 634
Db 612 GAGGTCGGACCGAGTACTGAGCAAAATGCTTCTCAGCTCTCTTTCGCTGCTCCCTAGC 671
QY 635 TTGCTTCAGAAAGTGTCTAATTTGGAGCAACTTTTTCGGTGAATGTTTTCGACCAATGG 694
Db 672 CTATTCAGAAAGTGTCTTTCGAGCCAGCTTTCGCGGCGCTTGAATTCCTACCAAGATGG 731
QY 695 CTAGATTTTCTGAGTCTACAGCGCGCTGAGGCGCTTGAAGGACCGGAGAACCAAGCC 754
Db 732 CTAGATTTTTCGCGCAAGTACTAGCGCGCTTTCGAGAGCTTCGAGAGACCGGAGAACTGAC 791
QY 755 GCCCCACAGCGCAGAAATCTTTGAGTT-CGCTGCGCTCGACTCTCTC--AGACCCCTC 810
Db 792 GCTCCACGCTGCAAGTCTCTGAGTTA CACTGCGCTGGAATCTTTCGCGAGACTCGC 851
QY 811 GGTCAACAGTCCCCTTGATTTGGCTAGAGAGGGATCCACTTGGCAATPACTTCGCCCCAGA 870
Db 852 GTCGTGAGTACTCTTTCGATTTGGCTAGAGAGGGACTCCAGTGGCAGTCTGCTGCTCAGA 911
QY 871 CCTAAATTTGGAGCTTAAGCCAGGCAAGTGTCTTTCGACCTTCGACGACAGGCTTTTCT 930
Db 912 -CTGGAAGTTAAACTCAAGGCCCAAGAAAGCTTTTAGACTCTGACGCGCCACTTTTCT 970
QY 931 CTTAGACAGCAGCTGTGGGGAGTGGAGCTGTGTGCCAGTAGCTTCAATCCCTCTGTGA 990
Db 971 CTTGAGCAGCAGCTGTGGGGAGTGGAGTTGCTGCCAGTAGCTTCAAGCTGCTTAGT 1030
QY 991 CTCTAAACCGGAACTTGGCTCTTTCGCGCTCTGCGCTCTTAAACATTCACGATAGACAA 1050
Db 1031 CTCCACCGAGAACTTGACTCTTCATCTCTGCGGCTCTGAGCGTTTCAGAGCTTAGGTAA 1090
QY 1051 TTTTCAGTGTGTTATCTTATTTGCTGAACCTTCTCTTCTGAGCTGCTTCTCTAGGCTAGA 1110
Db 1091 TTTCAAGGTAGTTCTCTATCTCTCTGAACCTTCTCTTCTGAGTACCTTCTCCAGTTAGG 1150
QY 1111 AGTCAGCTATCAGAAACAGTGTATGAAATAGCGAGTGTAGTTCGGCTTTCAGACACTAACCCC 1170
Db 1151 GCTCGCTGTTCAGAGCAGCTGTGGAGTGGCCAGTTTGTGGTTTTCGGAACACTAACCCC 1210
QY 1171 AGAGAGCAGTGTCTGAGAGAGGACCAATTTGTATCCCGCGCTGAGTGCAGAACTCAT 1230
Db 1211 AGAGAGCTGTCTTCTTCTGAAAGATGTTGTCTCATCTCTGCGCTTGGCGGAGAGATGTC 1270
QY 1231 TCCGCGCTCTGTCGAGCGGATGTCACCTCTTCTTTCAGGAAAGGCTTACAGAAATTCACCA 1290
Db 1271 GGCACCGCTTGAAGAGGTTGTCGCTCTCTCTTACAGAGGCTTTCGCGAAATTCACCA 1330
QY 1291 TCTTTCGATGAAACGCGCTGGAATTTCTTTCAACAGGCTTAAACAGGGGCAAGATTTACCCAC 1350
Db 1331 CTTTCGTATGAAACGCGCTAGAAATTTCT--CCAGGCTAAACAAAGGGCAAGAGTTTACCCAC 1387
QY 1351 CCCTGACAGGATTAATGGCTTACCACTGAGGAGGAAACAGAGCTTCTTCCGGATGGA 1410

```


Db 1388 CCCTGACCAAGATATGCTATCATAGCCTCGAGGAGGAAATAACTCTTCTCCGGATGGA 1447
Qy 1411 TCCAAAAACATGCGAGAGATAACCCAAACACAGTTGTTCTCTGCTGCGAGACATTCCTGG 1470
Db 1448 CCCAACAATTTGCACAGATACCCACGACACGCGGTGTCCTCTGCTGCAGACAGGCC --- 1504
Qy 1471 AAACACCAGGAATCCACTGAAGAAAATAAGAAATTAATACTACAGAGTTCCACTTGC 1530
Db 1505 -----GGAGCCCACTGAGAAAACCCAGAAATTTGGTGAATCAAGAAAGTTTCAC ---- 1551
Qy 1531 TTTGGAAGAGAGAGCCCTCTGAGGGCTGTCATCTAGTCAGATACCTATGGAAGAAAGGA 1590
Db 1552 -----AGAGCCCCAGGAGACGAGTCTGTTTGTGAATACCCGTGGAAGAAAGA 1600
Qy 1591 GCCTGGAGAGGCCGGAATAGTGTAGTTGATTACTCATACCTAGAAAGTGACCTTCCCAT 1650
Db 1601 ATGTGAGAGGACCACTAATGCACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGT 1660
Qy 1651 TTCTGCCAGACAGCTTGTAGTAACAACTGATAGATTAATTTTGGAGGTGCATCCAG 1710
Db 1661 TTCTACCAGACAGTTGTAGCAACAACTGATAGATTAATTTTGGAGGCGCCCCAG 1720
Qy 1711 TGACCTGGAACAAAGTTCTGATCAGAGAGTGAGGATTTGGATCAGGAAGCTGAGGATGA 1770
Db 1721 TGACTTGGAGCCAGCTCTGATTTCTGAAGTGAGGATTTGGGCGAGGAACTGAGGAGGA 1780
Qy 1771 TGGTTTTGATAGTAGCTCACTGTCTCAGACTCAGACCTTGAACAAGACCTCGAAGGGCT 1830
Db 1781 TGGCTTTGATAGCATGCTCCCTGTCTGATCAGAGCTGGAACAGGACTCGGAAGGCT 1840
Qy 1831 TCACCTTTGGAATCTTTCTGAGTGATAGATCCCTTAATATCCCCAGAACCTTTACAGAAC 1890
Db 1841 TCACCTTTGGAATCTTTCCACAGGTAGATCCCTTACAAACCCCAAACTTTACAGCCAC 1900
Qy 1891 AATTCAGACTCTGCCAGAAATTTCTGAGAGGCTTCTGATTCAGAGAGGATTTGTC 1950
Db 1901 GATTCAGAGGCTGCGAGAAATTTGCCCCAGAGACCCCATCAGATTCAGGAGCATCCTGGTC 1960
Qy 1951 TGGCAAGTCTGATCTAGAGAAATTCCTCCAGCTGGAAGCCTCTCTGAGACCCCTGAGCA 2010
Db 1961 TGGCAGCTGTGTTAGGA ---GCTGTGAGGAGGACCCCTTCGGGAGACCCCGACCA 2017
Qy 2011 TAGTTCTGGGAGAGAGATGATCTGGGAATCTAGTGCAGATGAAGCAGAGATCTCAAACT 2070
Db 2018 TAGTTCCGGGAGGAAGATGATCTGGAAACCGAGTGCAGATGAAGCAGAGATCTTAAAT 2077
Qy 2071 GTGGAATCATTTCTGTAATCTGATGACCCCTCAACCTTTTAAATTTAAGGCTCCTTT 2130
Db 2078 GTGGAATCTTTCTGTCTATCTGAGGACCCCTCAACCTTTTAAATTTAAGGCTCCTTT 2137
Qy 2131 TCAAAACATCAGGGGAAAATGAGAAAGGCTGTCGTGACTCAAAAGACCCCATCTGAGTCCAT 2190
Db 2138 TCNACCGTCAGGGAAGAAATGGAAGGCCGTGAGGACTCAAAAGCCCTCTTCTGAGGTAC 2197
Qy 2191 TGTGGCCATTTCTGAGTGTCAACCTTACTTTCTTGAAGTGAGTGACGTGTTGGGGAGCCA 2250
Db 2198 AGTGGCCCTCTCTGGCCATCATACCTTACTTTCTTGAAGGCCAGCTGTTAGAGAGCCA 2257
Qy 2251 AGAAAGTCAATGTCCAGACTCGGTACAGCGTGAGTCTTTCTGAGGAGAGACACACACA 2310
Db 2258 AGAAGATTAATTTGTCAGGCTGTGGGCTGGGCTCTTCTGAGGAGAGATACACCCA 2317
Qy 2311 TGTCAAAAGAAAAGGTAACCTTCTTGAAGAAAGTTACTGAGTATTAATAAGTGGTGA 2370
Db 2318 TATCAAGAGAAAAGGTAACCTTCTTGAAGAAAGTTACTGAGTATTAATAAGTGGTGA 2377
Qy 2371 TGAGGATCGCAAGAGCAATGGGAAGAAATTTGCAAGGAGTGGATCGAGTTCCAGAAACG 2430
Db 2378 TGAGGATCGCAAGAGCAATGGGAAGAAATTTGCAAGGAGTGGATCGAGTTCCAGAAACG 2437
Qy 2431 AATTCAGAAACAGAGATGCTATTTGGATATTTGTCATTTGACATTTGACACAGAGAAAGAT 2490
Db 2438 AATTCAGAAACAGAGATTTGCCATTTGGCTACTGTCTGGCCCTTTGAGCAACAGAGAAAT 2497

RESULT 5

US-10-296-115-224
; Sequence 224, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; FILE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 224
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-224

Query Match 28.2%; Score 828.6; DB 18; Length 883;
Best local Similarity 97.7%; Pred. No. 2.9e-237;
Matches 863; Conservative 0; Mismatches 14; Indels 6; Gaps 2;

Qy 1709 AGTGACCTGGAAAACAAGTTCTGATCCAGAAAGTGAGGATTTGGATGAGGAAGCTGAGGAT 1768
Db 1 AGTGACCTGGAAAACAAGTTCTGATCCAGAAAGTGAGGATTTGGATGAGGAAGCTGAGGAT 60
Qy 1769 GATGGTTTGTATGATGATAGTCACTGTCTGAGCTCAGACCTTGAACAAGACCTTGAAGG 1828
Db 61 GATGGTTTGTATGATGATAGTCACTGTCTGAGCTCAGACCTTGAACAAGACCTTGAAGG 120
Qy 1829 CTTCACTCTTGGAACTCTTCTGAGTGATAGTCCCTTATTAATCCCAAGACTTTTACAGCA 1888
Db 121 CTTCACTCTTGGAACTCTTCTGAGTGATAGTCCCTTATTAATCCCAAGACTTTTACAGCA 180
Qy 1889 ACAATTACAGACTGCTGCAGAAATTTGTTCTTGAAGAGCTTCTGATTCAGAGAAAGATTG 1948
Db 181 ACAATTACAGACTGCTGCAGAAATTTGTTCTTGAAGAGCTTCTGATTCAGAGAAAGATTG 240
Qy 1949 TCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAGCTTCTCTGAGACCCCTGAG 2008
Db 241 TCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAGCTTCTCTGAGACCCCTGAG 300
Qy 2009 CATAGTTCTGGGAGGAGAGATGATCTGGGAATCTAGTGCAGATGAAGCAGAGACTCTCAA 2068


```

RESULT 8
US-10-046-935-2040
; Sequence 2040, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguo
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secretist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046.935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2040
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8, 9, 14, 354, 527
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-2040

```

Query Match	9.0%;	Score 265.8;	DB 13;	Length 558;
Best Local Similarity	96.4%;	Pred. No. 2.5e-66;		
Matches 270;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;
Qy	2663	TTGGATCTAGTCTTAATTTTTTTTTTTTTCGCACATATCCCACTCAGAAACATTCAGGTTTGA	2722	
Db	3	TTGGGNNCTAGNTTAAATTTTTTTTTTTCGCACATATCCCACTCAGAAACATTCAGGTTTGA	62	
Qy	2723	AGCCAGGCCCTGATATGAAGGATGAACTAGTGTGATTTCTTAATCCTCCCTTTTTTGATTT	2782	
Db	63	AGCCAGGCCCTGATATGAAGGATGAACTAGTGTGATTTCTTAATCCTCCCTTTTTTGATTT	122	
Qy	2783	AGTTGGATGTGCTTTTAAATGTCCTTTGCCCTGCACTGAGGTGGAAAGGGGACCTTTTTTGAG	2842	
Db	123	AGTTGGATGTGCTTTTAAATGTCCTTTGCCCTGCACTGAGGTGGAAAGGGGACCTTTTTTGAG	182	
Qy	2843	TTGTCAATTTTGCACCTTTCAAAACCTATTTTCTTGGAAAAACAATATTTATAGGGCTTAAAG	2902	
Db	183	TTGTCAATTTTGCACCTTTCAAAACCTATTTTCTTGGAAAAACAATATTTATAGGGCTTAAAG	242	
Qy	2903	CCCAATTTTCATTTCTAAATCTAAATATATGTGTGCCTATCTG	2942	
Db	243	CCCAATTTTCATTTCTAAATCTAAATATATGTGTGCCTATCTG	282	

```

RESULT 9
US-10-146-502-2040
; Sequence 2040, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2040
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
;
; NAME/KEY: misc_feature
; LOCATION: 8, 9, 14, 354, 527
; OTHER INFORMATION: n = A, T, C or G
US-10-146-502-2040
```

	Query Match	9.0%;	Score 265.8;	DB 14;	Length 558;
	Best Local Similarity	96.4%;	Pred. No. 2.5e-68;		
	Matches 270;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;
Qy	2663	TTGGATCTAGTTAAATTTTTTTTTTTGGCAACATATCCCACTCAGAAACATTCAGGTTTGA	2722		
Db	3	TTGGGNNCTAGNTAAATTTTTTTTTTTGGCAACATATCCCACTCAGAAACATTCAGGTTTGA	62		
Qy	2723	AGCCAGCCCTGATAATGAAGGATGAACCTAGTGTGATTTCTTAATCTCTCCCTTTTTTGATTT	2782		
Db	63	AGCCAGCCCTGATAATGAAGGATGAACCTAGTGTGATTTCTTAATCTCTCCCTTTTTTGATTT	122		
Qy	2783	AGTTGGATGTGCTTTTAAATATGCTCTTTGGCTGCATGAGGTGAAAGGGGACCTTTTGTAG	2842		
Db	123	AGTTGGATGTGCTTTTAAATATGCTCTTTGGCTGCATGAGGTGAAAGGGGACCTTTTGTAG	182		
Qy	2843	TTGTCAATTTTGGCACTTTCAAAACTTAATTTTCTTTGGAAAAACAATATTTATAGGGCTTAAAG	2902		
Db	183	TTGTCAATTTTGGCACTTTCAAAACTTAATTTTCTTTGGAAAAACAATATTTATAGGGCTTAAAG	242		
Qy	2903	CCCAATTTTCATTTCTTAATCTAAATTAATGTGTGCCTATCTG	2942		
Db	243	CCCAATTTTCATTTCTTAATCTAAATTAATGTGTGCCTATCTG	282		

```

RESULT 10
US-09-920-300A-45/c
; Sequence 45, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Scaister, Heather
; TITLE OF INVENTION: COMPOSITIONS AND
; TITLE OF INVENTION: AND DIAGNOSIS O
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/92
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version
; SEQ ID NO 45
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 34_35, 39
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-45

```

	Query Match	6.8%	Score 199,	DB 9;	Length 706;
	Best Local Similarity	94.8%;	Pred. No. 3.4e-48;		
	Matches 217;	Conservative	0;	Mismatches 10;	Indels 2;
					Gaps 1;
Qy	2716	GGTTTGAGCCGCCCTGATAATGAAGGATGAAGTCTTAATCTT			
Db	697	GTTTGAAGCCGCCCTGATAATGAAGGATGAAGTCTTAATCTT			
Qy	2774	TTTTTGATTTAGTTGGATGTGCTTTTAAATGTCCTTTGCCCTGCATGAGTGGAAAGGGGAC			
Db	637	TTTTTGATTTAGTTGGATGTGCTTTTAAATGTCCTTTGCCCTGCATGAGTGGAAAGGGGAC			
Qy	2834	CTTTTTGAGTTGTCATTTTGGCATTTTCAAACTTTATTTTCTGGAAAACAATATTTATAG			
Db	577	CTTTTTGAGTTGTCATTTTGGCATTTTCAAACTTTATTTTCTGGAAAACAATATTTATAG			

	Query Match	6.8%	Score 199;	DB 22;	Length 706;
	Best Local Similarity	94.8%;	Pred. No. 3.4e-48;		
	Matches 217;	Conservative 0;	Mismatches 10;	Indels 2;	Gaps 1;
Qy	2716	GGTTTGAGCCACCCCTGATAATGAAGAGATGAAC	TAGTGTGATGTTCTTAATCCT	-CCCTT	2773
Db	697	GTTTGAAGCCACCCCTGATAATGAAGAGTGAAC	TAGTGTGATTCCTAATCCTCCCTT		638
Qy	2774	TTTTGATTTAGTTGGAATGTGCTTTTAAATGCTCTT	TGCTGCATGAGGTGGAAAGGGGAC		2833
Db	637	TTTTGATTTAGTTGGAATGTGCTTTTAAATGCTCTT	TGCTGCATGAGGTGGAAAGGGGAC		578

Qy 2834 CTTTGTGAGTGTGCTTTGACCTTCAAACTATTATTTCTTGAAAAACAATATTTATAG 2893
Db CTTTGTGAGTGTGCTTTGACCTTCAAACTATTATTTCTTGAAAAACAATATTTATAG 518
Qy 2894 GGCCTAAAGCCCATTTTCATTTCTTAATCTAAATATGTGTCCTATCTG 2942
Db 517 GGCCTAAAGCCCATTTTCATTTCTTAATCTAAATATGTGTCCTATCTG 469

RESULT 14

US-09-983-965-2096
; Sequence 2096, Application US/09983965

; Patent No. US20020137160A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 37-21(10297)C

; CURRENT APPLICATION NUMBER: US/09/983,965

; CURRENT FILING DATE: 2001-10-26

; PRIOR APPLICATION NUMBER: US 09/465,231

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: US 60/113,678

; PRIOR FILING DATE: 1998-12-17

; NUMBER OF SEQ ID NOS: 5912

; SEQ ID NO 2096

; LENGTH: 399

; TYPE: DNA

; ORGANISM: Bos taurus

; FEATURE:

; OTHER INFORMATION: Clone ID: 32-LIB3057-012-Q1-K1-H11

US-09-983-965-2096

Query Match 6.8%; Score 198.6; DB 9; Length 399;
Best Local Similarity 83.8%; Pred. No. 3e-48;
Matches 249; Conservative 0; Mismatches 44; Indels 4; Gaps 2;

Qy 2323 AAAGTAACTTCCTTGAAGAGTACTGAGTATTATATAGTGGTATGAGGATCCAA 2382
Db 102 ATAGGTAACCTTCCTTGAAGAGTACTGAGTATTATACATAAGCGGTGATGAGGATCGAA 161
Qy 2383 AGGACCATGGAGAAATTTGCAAGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 2442
Db 162 AGGACCATGGAGAAATTTGCAAGGATGGATGAGGTTCCAGAAACGAATTCAGAAAC 221
Qy 2443 AGAAGATGCTATTGGATATTGCTTTGACATTTGAAACACAGAGAAAGATGTTTAAATAGA-- 2500
Db 222 AGAAGAGCTATTGGATATTGCTTTGACATTTGAGCACAGAGAAAGATGTTTAAATAGA 281
Qy 2501 -CTCAGGGAACATGCTTCAAGGACTTAATGTTCTCAAGCAATGTTGAGTTG-GCAGCC 2558
Db 282 CTTCAAGCAACATGTTTCAAGGACTTAATGTTTTCGAGCAATGTTTAAAGATGATTCGAC 341
Qy 2559 TGTAAGTCTAGCTAGCATACACTACCTTACCTGAGAGGTTGCTTTTAAACACAA 2615
Db 342 AGCCTCTGGCCCTAGCATACACTACCTTACCTGAGAGGTTGCTTTTAAACACAA 398

RESULT 15

US-09-833-790-371

; Sequence 371, Application US/09833790

; Patent No. US20020068288A1

; GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.

; APPLICANT: Wang, Tongtong

; APPLICANT: Secrist, Heather

; APPLICANT: Mohamath, Raodoh

; APPLICANT: Indirias, Carol Y.

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121-512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 371
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-790-371

Query Match 6.4%; Score 187.4; DB 9; Length 628;

Best Local Similarity 99.5%; Pred. No. 9.5e-45;

Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2754 GTGATTTCTAATCCTCCCTTTTGTGATTTAGTTGGATGCTTTTAAATGTCCTTTGCCT 2813
Db 1 GTGATTTCTAATCCTCCCTTTTGTGATTTAGTTGGATGCTTTTAAATGTCCTTTGCCT 60
Qy 2814 GCATGAGGTGGAAGGGGACCTTTTGTGATTTGTCAATTTTGACCTTTCAAAACTTATTTTC 2873
Db 61 GCTTGAGGTGGAAGGGGACCTTTTGTGATTTGTCAATTTTGACCTTTCAAAACTTATTTTC 120
Qy 2874 TTGAAAAACAATATTTATAGGGCTTAAAGCCCAATTTTCATTTCTAATCTAAATATATGTGT 2933
Db 121 TTGAAAAACAATATTTATAGGGCTTAAAGCCCAATTTTCATTTCTAATCTAAATATATGTGT 180
Qy 2934 GCCTATCTG 2942
Db 181 GCCTATCTG 189

Search completed: September 16, 2005, 10:39:57

Job time : 1746.36 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 11:51:14 ; Search time 8782.62 Seconds
(without alignments)
12750.757 Million cell updates/sec

Title: US-10-650-482-1
Perfect score: 2942
Sequence: 1 attttggcttcgtccac.....aaattatgtgcttatctg 2942

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_hic.*
- 4: gb_est3.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_gest1.*
- 9: gb_gest2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605.4	54.6	3941	3	HSM805059
2	1364.4	46.4	2619	3	AK082957
3	1363.4	46.3	2731	3	AK049028
4	919.4	31.3	979	1	AL526219
5	868.8	29.5	897	5	BU162260
6	834.8	28.4	873	5	BQ221596
7	803	27.3	835	6	CD656897
8	777.2	26.4	1021	4	BM559080
9	768.2	26.1	775	6	CD642687
10	758	25.8	793	6	CA749002
11	755.6	25.7	795	6	CA426770
12	748.8	25.5	850	4	BG721970
13	736.4	25.0	875	5	BQ427697
14	734.6	25.0	838	7	CF594071
15	728.4	24.8	793	6	CB989883
16	719.4	24.5	792	7	CB538482
17	712.6	24.2	1047	2	BF797553
18	710	24.1	776	1	AI123720
19	709	24.1	721	5	BX510286
20	707.6	24.1	830	6	CB960555
21	699.8	23.8	764	6	CB990340
22	692	23.5	786	4	BG434646
23	685.8	23.3	784	6	CB960059
24	678	23.0	1637	3	CR595060

98	487.2	16.6	709	7	CF745488	CF745488 UI-M-GW0-
99	479.6	16.3	716	4	BI088423	BI088423 602852920
c 100	473.2	16.1	478	1	AI684846	AI684846 wa86c12.x
ALIGNMENTS						
RESULT 1						
HSM805059						
LOCUS	HSM805059 3941 bp mRNA linear HTC 22-SEP-2004					
DEFINITION	Homo sapiens mRNA; cDNA DKFZp6661186 (from clone DKFZp6661186).					
ACCESSION	AL833746					
VERSION	AL833746.1 GI:21734396					
KEYWORDS	HTC.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
	1 (bases 1 to 3941)					
AUTHORS	Bahr, A., Lauber, J., Mewes, H. W., Weil, B., Amid, C., Osanger, A.,					
	Fobo, G., Han, M. and Wiemann, S.					
	The German cDNA Consortium					
CONSRMT	Direct Submission					
TITLE	Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764					
JOURNAL	Neuberberg, GERMANY					
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer					
	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;					
	sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing					
CONSRMT	consortium of the German Genome Project.					
	This clone (DKFZp6661186) is available at the RZPD Deutsches					
	Reisourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.					
CONSRMT	Please contact RZPD for ordering:					
	http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp6661186					
	Further information about the clone and the sequencing project is					
FEATURES	available at http://mips.gsf.de/projects/cdna/ .					
	Location/Qualifiers					
	1..3941					
source	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="RZPD:DKFZp6661186"					
gene	/db_xref="taxon:9606"					
	/clone="DKFZp6661186"					
	/tissue_type="stomach"					
CDS	/clone_lib="666 (synonym: hato2). Vector pSport1; host					
	DH10B; sites NotI + SalI"					
	/dev_stage="adult"					
gene	/note="hypothetical protein, N-terminus truncated"					
	1..3941					
	/gene="DKFZp6661186"					
CDS	42..1225					
	/gene="DKFZp6661186"					
	/codon_start=1					
gene	/product="hypothetical protein"					
	/protein_id="CAH56240.1"					
	/db_xref="GI:52545825"					
CDS	/translation="ASKQQLPTPDQNGVHSLBEHSLRLMDPKHCEDNPTQFVPA					
	GDIPTQNTSETEEKIELTITVPLALEEESPGECPSEIPMEKEPGEGRISVVDYIS					
	LEGDLPTARSAPACSNKLIDYILGASDLSTSDPEGEDWDEAEADPGFSDSLSDS					
ORIGIN	DLEQDPEGLHWNFSFCSVPYNPNFTATIQTAAIRIVPEEPSEKDLGSKSLDENS					
	QSGSLPTPEHSGEEDWESSADEASLKLWNSFCNSDDPNPLNFAPQTSGENE					
	KGRDSTKTPSPSIVAISSCHTLKSKVOLLGSEPCDPSVQRDVLSGRRHTRKXK					
ORIGIN	VTFLEVEVTEYIISGEDRKGWBFARDGCRFQRQIETEDAIGYCLTFEHRERFNR					
	LQTCFKGLNVLKQC"					
	Query Match 54.6%; Score 1605.4; DB 3; Length 3941;					
ORIGIN	Best Local Similarity 99.9%; Pred. No. 0;					
	Matches 1617; Conservative 0; Mismatches 1; Indels 1; Gaps 1;					
	1324 GGCTAACAGGGGCAAGATTATACCCACCCCTGACACAGGATATGCTACCAACAGCTTGA 1383					
ORIGIN						
	1 GGCTAGCAAGGGGCAAGATTATACCCACCCCTGACACAGGATATGCTACCAACAGCTTGA 60					

Qy	2464	CTTGACATTTTGAACACAGAGAAAGATGTTTAATAGACTCCAGGGAACATGCTTCAAGG	2523
Ds	1141	CTTGACATTTTGAACACAGAGAAAGATGTTTAATAGACTCCAGGGAACATGCTTCAAGG	1200
Qy	2524	ACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCCCTAGCTAGCATACACTAC	2583
Ds	1201	ACTTAATGTTCTCAAGCAATGTTGAGTTGGCAGCTGTAGTCCCTAGCTAGCATACACTAC	1260
Qy	2584	CTCTTACCTGAGAGGTGCTTTTAAACAACTTTGGCAGCTGTCTTTTGACATTTTTT	2643
Ds	1261	CTCTTACCTGAGAGGTGCTTTTAAACAACTTTGGCAGCTGTCTTTTGACATTTTTT	1319
Qy	2644	TTTTTAGGAGAAATGTAACCTGGATCTAGTTTAAATTTTTTTTTTGGCAATATCCCACT	2703
Ds	1320	TTTTTAGGAGAAATGTAACCTGGATCTAGTTTAAATTTTTTTTTTGGCAATATCCCACT	1379
Qy	2704	CAGAAACATTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTA	2763
Ds	1380	CAGAAACATTCAGGTTTGAAGCCAGCCCTGATATGAAGATGAACCTAGTGTGATTTCTA	1439
Qy	2764	ATCTCTCCTTTTGTAGTTAGTTGGATGTGCTTTTAAATGTCCTTCCCTGCATGAGGTG	2823
Ds	1440	ATCTCTCCTTTTGTAGTTAGTTGGATGTGCTTTTAAATGTCCTTCCCTGCATGAGGTG	1499
Qy	2824	GAAGGGGACCTTTTTCAGTTGTCATTTTGCACATTTTCAAACTTATTTCTTGGAAACA	2883
Ds	1500	GAAGGGGACCTTTTTCAGTTGTCATTTTGCACATTTTCAAACTTATTTCTTGGAAACA	1559
Qy	2884	ATATTTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATGATGTGCTCATCTG	2942
Ds	1560	ATATTTATAGGCTTAAAGCCCATTTTCATTTCTAACTAAATGATGTGCTCATCTG	1618
RESULT 2			
LOCUS	AK082957	2619 bp mRNA linear HTC 03-APR-2004	
DEFINITION	Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone: C530022L24 product: hypothetical protein, full insert sequence.		
ACCESSION	AK082957		
VERSION	AK082957.1	GI:26350136	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 Carninci, P. and Hayashizaki, Y.		
AUTHORS	1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		

REFERENCE	
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
REFERENCE	Nature 409, 685-690 (2001)
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2619)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, K., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
	Please visit our web site for further details.
	URL: http://genome.gsc.riken.jp/
	URL: http://fantom.gsc.riken.jp/
FEATURES	Location/Qualifiers
source	1..2619
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM DB:C530022L24"
	/db_xref="taxon:10090"
	/clone="C530022L24"
	/tissue_type="spinal cord"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="12 days embryo"
	418..2511
	/note="unnamed protein product; hypothetical protein (evidence: ProCrest)
	putative"
	/codon_start=1
	/protein_id="BAC38708.1"
	/db_xref="GI:26350137"
	/translation="METGTHARKRPGLRGSWFLPPLRRSHACSSPEPPSSRQNP
	GNALPERTRYTKLLSQLALLPSLFQKLLIWSQLFGLIPTRLDPAASYALRA
	LKREESAPTVQKSSLRDSDLVSSLDLWEEGLQWCSSDLKLAQKQRA
	LDSPAFTFLQGLQWGLLPSSLQAGLVSHRELSSSGPLSVQPLKPVVSYLLN
	PSYLDLPQLQRCQSSAGGQVGFRTITPESCYLSDGCHQIPQIARMSATWRR
	PLSTGLPEIHLRMKRLFLQANKQSLPTPDQNGVHSLSEHNLLRMDPHCTD
	NPAQVSPADPEKPELVIOEVSPQSSSLFCELPEVECEEDHTNATDLS
	RESLVPVTRPVCSKLIIDYILGGASDLSESDSESDGEPEDGDSGLSES
	DVEQDEGLHWNPSHVDYFNPQNTATQTAARAPRDPSGTSWGSQCGVQK
	EGPLPTPHSSGEEDDWEPSDAENLKLWNSFCHSDPYNLLNPAFPQPSGKWK
	GRQDSKASSEATVAFSGHTLLSCAKLLESDQNCPCGGLGALAGERYTHIKKV
	TFLEVTYVIGDEDRKGFWEFARDGCRFKRIQETEVIGCYCLAFEHKQVNL
	RIESKDLLSYNVKK"
ORIGIN	

Query Match		46.4%;	Score 1364.4;	DB 3;	Length 2619;
Best Local Similarity		73.9%;	Pred. No. 0;		
Matches 1900;		Conservative	0;	Mismatches 611;	Indels 59; Gaps 11;
Qy	96	CCATGCGCGAGCGCTGGGGGACCTCTATCCATGGTGTGAAGCGTCGAGCCGACATAGG	155		
Db	107	CGGTGCGCGCGCGAGGAGGGGTCTCTCTATGGTGGAGCGATCTCACACGCGCTAGG	166		
Qy	156	GAACCTCTCTCCC-CGCCAGGATGGAGTGCATCAGTCGCGGCTATGTCGGCGGCTGT	214		
Db	167	ACGTCTCTCTCCCTAGCGGATGGACCTAACCGCGTCGCCACCGCTTGGCGGGGCTC	226		
Qy	215	TCCTCCCTGTGTCTGCGCGCGCGCTGCGCATTCGCTCCCTCTGTGTGCTTTCTGTGG	274		
Db	227	TGGGCGGTCCGGTGCAGCACTGGTTGCGGAAGCGCGCTCTCTGGGCTCTCTGCGG	286		
Qy	275	CTCGAAGATCGGCTTGGAGCGACGACCGCAACCGCTGGGCAAGCGCGAGCTCTGTAGGCT	334		
Db	287	CGCGGAATCGAATGTCAGTATCCACCTCCGTGGCTGGCAAGGGCGAGACTGTGTAGACC	346		
Qy	335	TCCTCCGNAATCCGTCGACCTCCAGCCGCTGAGCGCGCGGCCCTACTCGAGAGACTGTC	394		
Db	347	TCGATCCAGCTCGCTGACCGCGCTGAGCTCTGTCTCTCTCTGTCTGAGAAGCGGCC	406		
Qy	395	AAGAAAGAGATGGAGCGGCGGACAGCGGATCGCGAAACGCGCTTGGCCCTCGGGCG	454		
Db	407	AAG-GAAAGGAGATGGAGACAGGAACGACAGGGCGCGGAAGCGGCTTGGCCCTCGGCTG	465		
Qy	455	GGCTTCGGTTCGTGGCCACCTTTTTCCTCGGGATCGCAAGCAGGCTCTTCTAAGTTC	514		
Db	466	GGCTCTCGTTCGGCTGCCCC---TTCTTTCGGGATCGCACGCGCTGCTCTTCGGAGTTC	522		
Qy	515	CGGAGCCTCTTGGCGCGGAAACTCCGGGAAACCCCACTGCTTCTCTGCGCCAGGCC	574		
Db	523	CCGCGCGCTCTCTCGACAAATTCGCGGAA-----CTCGCTCTCGCC	567		
Qy	575	GAGACTCGGGTCAGTTACTGACGAAACTGCTCTCCAGCTCTTGGCGCGCTCCCGGA	634		
Db	568	GAGCGTCGACCGAGTACTGACCCAAATGCTTCTCAGCTCTTGGCCCTGCTCCCTAGC	627		
Qy	635	TGCTTCAGAAAGGTGCTAAATTTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACAGATGG	694		
Db	628	CTAATTCAGAAAGTCTGCTTTTGGAGCAGCTTTTTCGGGGCGCTGATTCCTTACCAGATGG	687		
Qy	695	CTAGATTTTGTGAGTCTACAGCGCCTGAGAGCCCTGAGGGACGGGAAACACAGCC	754		
Db	688	CTAGATTTTTCGCAAGTCTACAGCGCCTGAGAGCTTTGAGAGACGGGAAATCTGCCC	747		
Qy	755	GCCCCCAGCGCGAGAAATCTTTGAGTTTCGCTGCGAGCTCGACTCTCTCAGACCCCTCGGTC	814		
Db	748	GCTCCACGGTGCAGAACTCTGAGTTCACTGGCGCTGGACTCTTCGGAAGACTTGGTT	807		
Qy	815	ACGAGTCCCCTTGATTTGGCTAGAGAGGGGATTCATGCGGAATACTTCGCCCCGACACTA	874		
Db	808	GTCAGTCTCTTGATTTGGCTAGAGAGGGAATCCAGTGGCAGTGTCTGCTCCTCAGACCTG	867		
Qy	875	AAATTTGAGCTTAAGGCCAAGGAGTCTTTGGACCTCGACACAGCGCTTTTCTCTTA	934		
Db	868	GAGTTAAACATCAAGGCCCAGAAAGAGCTTTAGACTCTGACGCGCCCACTTTTCTCTG	927		
Qy	935	GAGCAGAGCTGTGGGAGTGGAGCTGTTGGCCGAGTAGCTTCAATCCCGCTCTGTACTCT	994		
Db	928	GAGCAGAGCTGTGGGAGTGGAGTTGCTGCCAGTAGCTTCAAGCTGGTCTAGTCTCC	987		
Qy	995	AACGGGAACCTTGGCTCTTGGCCCTCTGGGCTCTTAACATTCACAGCATAGACAAATTC	1054		
Db	988	CACCGAGAACTTGAATCTTCTCATCTCTGGGCTCTGAGCGTTTCAGAGCTTAGGTAATTC	1047		
Qy	1055	AGTGTGTATCCTATTTGCTGAACCTTCTTACCTGGAGCTGCTTCTTAGGCTAGAGTTC	1114		
Db	1048	AAGGTAGTTTCTATCTCTGAAACCTTCTTACCTGGAGTACTACCTTCCCGAGTTAGGGCTG	1107		
Qy	1115	AGCTATCAGAAACGTGATGGAATAGCGAGGTAGTCGGCTTCCAGACACTAACCCCGAG	1174		

Db	1108	CGCTGTCCAGACGAGCGCTGGAGGTGGCAGTTTGTGGTTTCCGAACTAACCCGAGAG	1167		
Qy	1175	AGCAGCTGCTGAGAGAGGACCATTTGTTCATCCCGAGCGCTGAGTCGAGAACTCATTTCCG	1234		
Db	1168	AGCTGCTATCTTTCTGAAGATGGTTGTACCTCAGCCGTTGGCGGCGAGAGATGTCGGCA	1227		
Qy	1235	GCCTGTCGGCAGGATGTCACCTCTTTCTACGGAAGGCTTACAGAAATTCACACTCTT	1294		
Db	1228	ACGCTTGGAGAGGTGTCGCTCTCTCTACAGAGGCTTGCAGGAAATCCACCACTT	1287		
Qy	1295	CGCATGAAACCGGCTGGAAATTCCTTCAACAGGCTAAAGGGGCAAGATTTACCCACCCCT	1354		
Db	1288	CGTATGAAACCGGCTAGAAATTCCT---CCAGGCTAAACAAAGGGCAAGAGTTACCCACCCCT	1344		
Qy	1355	GACCAGATTAATGGCTACACAGCCTGGAGAGGACACAGCTTCTCCGAGTGGATGCCA	1414		
Db	1345	GACCAAGATTAATGGCTATATAGCTTGGAGGAGAAATTAACCTTCTCCGAGTGGAGCCCA	1404		
Qy	1415	AAACACTGACAGATTAACCCAAACACAGTTTGTTCCTGCTCGAGACATTTCTCTGGAAC	1474		
Db	1405	CAACATTCACAGATTAACCCAGCAGGCGGTGTCCTGCTGCGAGACAGGCC-----	1457		
Qy	1475	ACCCAGGAATCCACTGAAGAAATAAGAAATTAATACTACAGAGTTCCACTTGTCTTGT	1534		
Db	1458	-----GGAGCCCACTGAGAAATAAACAGAAATTTGGTATTCAAGAAAGTTTCAC-----	1504		
Qy	1535	GAAGAGAGAGCCCTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGAAGAGGAGCCT	1594		
Db	1505	-----AGAGCCCCCAGGGAAGCAGTCTGTTTGTGAAATACCCGTTGGAAGAAAGATGT	1557		
Qy	1595	GGAGAGGCGCAATTAAGTGTAGTTGATTACTCATACCTAGAAGGTGACCTTCCCATTTCT	1654		
Db	1558	GAAGAGAGCCACACTAATGCAACTGACCTCTCAGATAGAGAGAGAGCCCTTCTGTCTTCT	1617		
Qy	1655	GCCAGACCACTTGTAGTAAACAACTGATAGATTAATATTTTGGAGGTGCAATCAGGTGAC	1714		
Db	1618	ACCAGACCACTTGTAGCAACAACTGATAGATTAATATTTTGGAGGCGCCCCAGGTGAC	1677		
Qy	1715	CTGGMAACAAGTCTGATCCAGAGGTGAGGATGGGATCAGGAAGCTGAGGATGATGGT	1774		
Db	1678	TTGGAGCCAGCTCTGATCTGAAAGTGAGGATTTGGGGCGAGGAACCTGAGGAGCATGGC	1737		
Qy	1775	TTTGATAGTATGACTCACCTGTGAGACTCAGACCTTGAAACAAAGACCCCTGAAGGCTTCAC	1834		
Db	1738	TTTGATAGCATGGCTCCCTGCTGTAATCAGACGTGGAACAGGACTCGGAAGGCTTTCAC	1797		
Qy	1835	CTTTGGAACCTCTTTCTGAGGTAGATCCTTATAATCCCGAGAACTTTTACAGCAACAAT	1894		
Db	1798	CTTTGGAACCTCTTTCCACAGTGTAGATCTCTTACAACCCGCCCAAACTTTTACAGCCACGATT	1857		
Qy	1895	CAGACTGCTGCCAGAAATTTGCTGAGAGGCTTCTGATTCAGAGAGGATTTGCTCTGGC	1954		
Db	1858	CAGACGCTGCCAGAAATTTGCCCCAGAGACCCATCAGATTCAGGGACATCTGTGCTGGC	1917		
Qy	1955	AAGTCTCATCTAGAGAAATTCCTCCAGTCTGGAAGCTTCTCCTGAGACCCCTGAGCATAGT	2014		
Db	1918	AGCTGTGTTAGGGA---GCTGTGAGAGGAGACCCCTTCCGGAGACCCCGACCATAGT	1974		
Qy	2015	TCGGGGAGGAAGATGATCGGGAATCTAGTGGAGATGAAGCAGAGAGTCTCAAACTGTGG	2074		
Db	1975	TCGGGGAGGAAGATGATCGGGAACCGAGTGCAGATGAAGCAGAGAAATCTTAAATTTGG	2034		
Qy	2075	AACTCATTTCTGAATTTCTGATGCCCTTACAACCTTTAAATTTTAAGGCTCTCTTTTCAA	2134		
Db	2035	AACTCTTTCTGTCAATCTGAGGACCCCTACAACCTTTAAATTTTAAGGCTCTCTTTCAA	2094		
Qy	2135	ACATCAGGGGAAATGAGAAAGGCTGTCTGTCATCAAAAGACCCCATCTGAGTCCATTTGTG	2194		
Db	2095	CCGTACGGGAAGAAATGGAAGGCGCTCAGGACTCAAAAGGCTCTTCTGAGGCTACAGTG	2154		
Qy	2195	GCCATTTCTGAGTGTACACCTTACTTTCTTTGAAGGTGAGCTGTTGGGGAGCAAGAA	2254		

Db	2155	GCCTTCTCTGGCCATCATACCTTACTTTCTTTGTAAGGCCACGCTGTTAGAGACCCAGAA	22121
Qy	2255	AGTGAATGTCCGACTCGGTACACGCTGACGTTCTTTCTGGAGGAAAGACACACATGTC	2314
Db	2215	GATAATTTGCCAGGCTGTGGGCTGGGTGAGGCTCTTTGCTGGAGAAAGATACACCCATATC	2274
Qy	2315	AAAGAAAAAAGGTAACTTCCTTTGAAGAAGTTACTGAGTATTATATTAAGTGGTGATGAG	2374
Db	2275	AAGAGAAAAAAGGTAACTTCCTTCTGGAAGAAGTTACTGAGTATTATATTAAGTGGTGATGAG	2334
Qy	2375	GATCGCAAGGACCATGTGGGAAGAATTTGCAAGGGATGGATGCGAGTTTCCAGAAACGAATT	2434
Db	2335	GATCGCAAGGACCATGTGGGAAGAATTTGCAAGGGATGGATGCGAGTTTCCAGAAACGAATT	2394
Qy	2435	CAAGAACACGAAGATGCTATTGGATATTGCTTGACATTTGAAACACAGAGAAAGATGTTT	2494
Db	2395	CAAGAACACGAAGTTGCCATTGGCTACTGCTTGGCCCTTTGAGCACAGAGAAAAATGTTT	2454
Qy	2495	AATAGACTCAGGGAACATGCTTCCTTCAAGGACTTAAATGTTCTCAAGCAATGTTGAGTTG-G	2553
Db	2455	AATAGACT-----GAGGATCGAGTCAAGGACTTACTGTTGTACAGCAATGTTTGAAGTG	2510
Qy	2554	CAGCCTGTAGTCTCTAGCTACATACACTCTCTTACTCTGAGAGGTGCTCTTTTAAAAACA	2613
Db	2511	AACAGCCCTGCAAGCCGTCGCCACTCTGCTCTTACTTGAGA-GTTCCTCTAAAAACA	2569
Qy	2614	AATCTTGGCAGCTGTCTCTTTGACATTTTTTTTTTTTAGAGGAAATGTAACT	2663
Db	2570	AACACTGGCAGCTGTCTCTTGGACATGTTTTTTTAAAGAAACAACCTTGTTATCT	2619
RESULT 3			
AK049028			
LOCUS			
DEFINITION			
ACCSSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

		Matches 1957; Conservative 0; Mismatches 641; Indels 79; Gaps 13;		
Qy	123	TATCATGTGTGAAGCGTCGAGCGGCTAGGAACTCTCTTCCC-GCCAGAGTGA	181	
Db	2	TCTCTATGTGAGCGATCTCACGCGCTAGGACGTCTCTCTTCCCTAGCCGGATGGAC	61	
Qy	182	GTCCGATCAGTCGCGCCCTATTGCGCGGCTGTCTTCCCTGTGTCTGCGCCCGCTGC	241	
Db	62	CTAACCGCGGTGCGCACCGCTTGCGCGGCTCTGGCGCTCGCGTGCAGCACGCTTGC	121	
Qy	242	CGCATTCGTCCTCTGTGGCTTTTCTGTGCTCGAAGATCGGCTGGAGCAGCGACG	301	
Db	122	GGAAAGCCCGCTCTCTGCGCTCTCTGCGGCTCTGCGCGCGGGAATCGACATGCCACT	181	
Qy	302	CCACCGCTGGGCAAGCGCGAGACTCTGTAGGCTTCCCTCGGAATCCCGTCGACCTCCAGCC	361	
Db	182	CCGTGGCTGGGCAAGCGGAGACTGTGTAGACTTCGGATTCAGCTCGCTGTGAGCGCGCT	241	
Qy	362	GCTGAGCGCGCGGCTCTACCTGAGAGACTGTCAAGAAAGAGAGATGGAGCCCGGGACA	421	
Db	242	GAGCTCTGTCTCTCTCTGTCTGAGAAGCGCCAG-GAAGGAGATGGACAGGAAAGC	300	
Qy	422	GGCGATCGCGGAACGCGTGTGGCTTCGCGGGGCTTCGCGTTCGCGCCACCTTTTTC	481	
Db	301	CACAGGCGCCGGAAGCGGCTGCGCTCGGCTGGGCTCTGTGTTCCGGCTGCC---TTC	357	
Qy	482	CCTCGCGCATCGAAGCAGGCTCTTTAAGTTCCGACGCTCTTGGCCCGGAAACTCC	541	
Db	358	CTTCGGCGATCGACGCTGTCTTCGGAGTTCCCGCGCTTCCTCTCGAAGAAATCCC	417	
Qy	542	GGGAACCCACACTGCTCTCTGCCAGCCCGAGACTCGGGTCAGTTACTGGACGAAA	601	
Db	418	GGGAACCTCGCTCT-----GCCCGAGCTCGGACCACTGTCGACCAA	462	
Qy	602	CTGCTCTCCAGCTCTTTCGCGCGCTCCCGGATTCGTTTCAGAGGTGCTAAATTGGAGC	661	
Db	463	TTGCTTTCTCAGCTCTCTTGCCTGCTCCTAGCTTAITCCAGAAAGTCTGCTTTGGAGC	522	
Qy	662	CAACTTTTCGGTGGAAATGTTCCGACAGATGGCTAGATTTTGTGGAGTCTACAGCGCC	721	
Db	523	CAGCTTTTCGGGGGCTGATTCCTACAGATGGCTAGATTTTTCGCGAAGTTACAGCGCC	582	
Qy	722	CTGAGAGCCCTGAAGGGAGGGAGAAACAGCGCCGCCACAGCGCGAATACTTTTGAGT	781	
Db	583	CTGAGAGCTTTGAGAGGAGCGGAGAAATCTGCCGTCCACAGTGCAGAACTCTCTGAGT	642	
Qy	782	TGCTGAGCTCGACTCTCAGACCCCTCGGTGACAGTCCCTCTGATTTGCTGAGAGAG	841	
Db	643	TCACTGCGGCTGGAATCTTCGGAAGACTTGGTTGTCTGCTCTCTTTGATTTGGCTAGAGG	702	
Qy	842	GGGATCCACTGCAATACTCGCCCCCAGACCTAAATTTGGAGCTTAAGGCCAAGGAGT	901	
Db	703	GCACTGCAGTGGCAGTGTCTGCTCTCAGACTGGAGTTAAAACTCAAGGCCCAAGAAAGA	762	
Qy	902	GCTTTGGACCTTGCAAGCAGGCTTTTCTTTAGAGCAGAGCTGTGGGAGTGGAGCTG	961	
Db	763	GCTTTAGACTCTGAGCGCCACATTTCTCTGAGCAGCAGCTGTGGGAGTGGAGTTG	822	
Qy	962	TGCCCCAGTAGCTTCAATCCCGTCTGTACTCTAAACCGGGAATTTGGCTCTTCCGCTCT	1021	
Db	823	CTGCCCACTAGCTTCAAGCTGGTCTAGTCTCCCAACCAGAACTTGAATCTTCTCATCTCT	882	
Qy	1022	GGGCTCTAAACATTTCAAGCATAGACAAATTTCAAGTGTGATCTTAATTTGCTCAACCT	1081	
Db	883	GGGCTCTGAGGCTTCAGAGCTTAGGTAAATTTCAAGGTAGTTTCTATCTCTGAAACCT	942	
Qy	1082	TCCTACCTGGACTGTCTTCTTAGGCTAGAAGTCAAGCTATCAGAACAGTGTGGAATAGC	1141	
Db	943	TCCTACCTGGACTACTTTCCCGAGTTAGGCTGCGCTGTCTCAGACAGCGCTGGAGGTGC	1002	
Qy	1142	GAGTAGTCGGTTCAGACACTAACCCAGAGAGAGCTGCTGCTCAGAGAGGACCATGT	1201	
Db	1003	CAGTTTGTGGTTCGGAACACTAACCCAGAGAGCTGCTATCTTTCTGAAGATGGTTGT	1062	

Qy	1202	CATCCCGAGCGCTGAGTCAGGAACACTCATTTCCGCGCTCGTGGCAGGATGTCCACCTCTT	1261	
Db	1063	CACCTCAGCGTTGCGGCGAGAGATGTCCGCAACCGCTGGAGAAGGTGTCCGCTCTC	1122	
Qy	1262	TCTACGGAAGCCCTACAGAAATTCACCATTTGCGATGAACCGGTGGAAATTCCTTCAA	1321	
Db	1123	TCTACGAAGGCTTCCGGAATTCACCACTTCGTATGAACCGCTAGAATTCCT---C	1179	
Qy	1322	CAGCTTAACAGGGGCAAGATTTACCCACCCCTGACCAGGATTAATGCTTACCACGCTG	1381	
Db	1180	CAGCTTAACAAAGGCAAGATTTACCCACCCCTGACCAGATTAATGCTTATCATAGCTG	1239	
Qy	1382	GAGGAGAAACACAGCCTTCTCCGATGGATCCAAAAACATCTGCAGAGATAACCCAAACAG	1441	
Db	1240	GAGGAGAAACATTAACCTTCTCCGATGGACCCACACATTTGCACAGATAACCCAGCAG	1299	
Qy	1442	TTTGTTCCTCTGCTGGAGACATTCCTGGAAACACCCAGAAATCCACTGAAGAAAAATA	1501	
Db	1300	GCGTGTCCCTGCTGCAGACAGGCC-----GGAGCCCACTGAGAAAAAACCG	1347	
Qy	1502	GAATTATTAACACAGAGTTCCACTTGTCTTTGGAAGAGAGAGCCCTTCTGAGGCTGT	1561	
Db	1348	GAATTGTGTGATTCAAGAAATTTTAC-----AGAGCCCCCAGGGAAGCAGT	1392	
Qy	1562	CCATCTAGTCAGATACCTATGGAAAGGAGCCCTGGAGAGGCCGAATAAGTGTAGTTGAT	1621	
Db	1393	CTGTTTGTGAAATTAACCGTGGAAAGAAATGTGAGAGGACACACTAATGCACATGAC	1452	
Qy	1622	TACTCATACCTAGAAGTGACCTTCCATTTCTGCCAGACCACTGTGTAGTAACAAATG	1681	
Db	1453	CTCTCAGATAGAGAGAGAGCTTCTCTGTTTCTACAGACCACTTGTAGCAACAACTG	1512	
Qy	1682	ATAGATTATATTTTGGAGGTGCATCCAGTGACCTCGAACAAGTTCTGTATCCAGAGT	1741	
Db	1513	ATAGATTATATTTTGGAGGCGCCCGCAGTGACTTGGAAAGCCAGCTCTGATTTCTGAAGT	1572	
Qy	1742	GAGGATTTGGAGTACGGAAGCTGAGGATGATGTTTGTATAGTATAGCTCACTGTCTCAGAC	1801	
Db	1573	GAGGATTTGGGCGAGGAACTGAGGACATGGCTTTGATAGCATGGCTCCCTGTCTGAA	1632	
Qy	1802	TCAGACCTTGAAACAAGACCTGAAAGGCTTCACTTTTGGAACTCTTTTCTCAGTGTAGAT	1861	
Db	1633	TCAGACGTGAAACAGGACTCGGAAGGCTTTCAGCTTTGGAACTCTTTCCACAGTGTAGAT	1692	
Qy	1862	CCTTAATATCCCAAGAACTTTACAGCAAAATTCAGACTGCTGCCAGAAATTTCTCTGAA	1921	
Db	1693	CCTTACAAACCCCAAACTTTACAGCCACGATTCAGACGCTGCCAGAAATTTGCCCCAGA	1752	
Qy	1922	GAGCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATTTCTCCACG	1981	
Db	1753	GACCATCAGATTCAGGGACATCTGTGTCTGGCAGCTGTGTGTAGGGA---GCTGTGAG	1809	
Qy	1982	TCTGGAAGCCTTCTGAGACCCCTGAGCAATGATTTCTGGGAGGAAGATGATCGGAATCT	2041	
Db	1810	GAGGAGCCCTTCCGGAGACCCCGGACCATAGTTTCCGGGAGGAAGATGATCGGAACCG	1869	
Qy	2042	AGTCAGATGAAGCAGAGAGTCTCAAACTGTGAACTCATTTCTGTAATTTCTGATGACCC	2101	
Db	1870	AGTCAGATGAAGCAGAGATCTTAAATGTGGAATCTTTCTGTCTCATTTCTGAGGACCC	1929	
Qy	2102	TACAAACCTTTAAATTTTAAAGGCTCTTTTCAAAACATCAGGGGAAATAGAAAGGCTGT	2161	
Db	1930	TACAACTTTTAAATTTTAAAGGCTCTTTTCAACCGTCAGGGAAGAAATGGAAGGCGGT	1989	
Qy	2162	CGTGACTCAAAGACCCCATCTGAGTCCATTTGTGGCCATTTTCTGAGTGTCTACCTTACTT	2221	
Db	1990	CAGGACTCAAAGGCTCTTCTGAGGCTTACAGTGGCTTCTCTGGCCATCATACCTTACTT	2049	
Qy	2222	TCTTGTAAAGTGTGAGTGTGGGAGCCAGAAAGTGAATGTCCAGACTCGGTACAGCGT	2281	
Db	2050	TCTTGTAAAGGCCAGCTGTTTGTAGAGGCCAAGAAAGATAATTTGTCCAGCTGTGGGCTGGT	2109	

```
Qy 2282 GACGTTCTTTCTGGAGGAGACACACATGTCACAAAGAAAAAGTAACCTTCTCTGAA 2341
Db 2110 GAGGCTCTTGTGAGAGAAAGATACACCCATATCAAGAGAAAAAGTAACCTTCTCTGAA 2169
Qy 2342 GAAGTTACTAGTATTATATAAGTGGTGATGAGATGCAAGAACCATGCGGAAGAATTT 2401
Db 2170 GAAGTTACTAGTATTATATAAGTGGTGATGAGATGCAAGAACCATGCGGAAGAATTT 2229
Qy 2402 GCAAGGATGATGATGAGTTCAGAAACGAATTCAGAAACAGAAAGATGCTATTGGATAT 2461
Db 2230 GCAAGGATGATGATGAGTTCAGAAACGAATTCAGAAACAGAAAGTTCGCTTGGCTAC 2289
Qy 2462 TGCTTGACATTTGAACACAGAGAAAGATGTTTAATAGACTCCAGGGAACATGCTTCAAA 2521
Db 2290 TGCTTGGCTTTGAGCACAGAGAAAAAATGTTTAATAGACT---GAGGATCGAGTCAAA 2345
Qy 2522 GGACTTAATGTTCTCAAGCAATGTTGATGGGAGCCCTGTAGTCTTAGCTAGCATACAT 2581
Db 2346 GGACTTAATGTTCTCAAGCAATGTTTAAGAGTGA-ACAGCCTGCAACCCCGTCCCACTCT 2404
Qy 2582 ACCTCTTACCTGAGAGGTGCTTTTAAACAAATCTTGGCAGCTGCTTTTGACATTTT 2641
Db 2405 GTCTCTTACTTGAGA-GTTCCTTAAACAAACACTGGCAGCTGCTTTGGACATGTT 2463
Qy 2642 TTTTGTAGAGAAATGAACTTGATCTAGTCTAGTCTTAAATTTTGTGCAACATATCCCA 2701
Db 2464 TTTAAAGAAACAACTTGATCTAGAGATGCAAGTTGATTTTGGTAAATGTCCTCA 2523
Qy 2702 CTCAGAAACATTCAGTTTGAAGCCAGCCCTGATTAAGAGGATGAAGTGTGATTTTC 2761
Db 2524 TTAGAA-----ACACCACTCCGATTAAGAAAT-----CTCTTATCTG 2563
Qy 2762 TAATCTCTCTTTTGTAGTTAGTTGATGCTTTT 2798
Db 2564 TAATCTCTCTTTTGTAGTTAGTTGATGCTTTT 2600
```

RESULT 4

```
AL526219 979 bp mRNA linear EST 24-MAR-2004
LOCUS AL526219 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC016Y122 5-PRIME, mRNA sequence.
ACCESSION AL526219
VERSION AL526219.3 GI:45701376
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 979)
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 13, 2001 this sequence version replaced gi:31064080.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
858.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0DC016BE11QPI&c=858.r.

FEATURES

source

```
1. 979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC016Y122"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
```

/clone lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

```
Query Match 31.3%; Score 919.4; DB 1; Length 979;
Best Local Similarity 96.2%; Pred. No. 1.5e-247;
Matches 935; Conservative 14; Mismatches 22; Indels 1; Gaps 1;

Qy 55 GCATTGCTCAGGGGCTTTTCAACCCCTCTGTCACTCGGAAAAACCATCGCCAGGCGCTGGG 114
Db 8 GCATTGCTCAGGGGCTTTTCAACCCCTCTGTCACTCGGAAAAACCATCGCCAGGCGCTGGG 67
Qy 115 GGACTCTCTATCATGCTGTGAAGCGTTCGAGCGGACTAGGGAACCTCTTCCCGCCAG 174
Db 68 GGACTCTCTATCATGCTGTGAAGCGTTCGAGCGGACTAGGGAACCTCTTCCCGCCAG 127
Qy 175 GATGGAAGTCGCATCACTCGCCGCTATTTCGCGGGGCTGTTTCTTCCCTGTGTTCTGCGC 234
Db 128 GATGGAAGTCGCATCACTCGCCGCTATTTCGCGGGGCTGTTTCTTCCCTGTGTTCTGCGC 187
Qy 235 CCCTGCGCGATTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTTGGAGC 294
Db 188 CCCTGCGCGATTCGCTGCGCTCTGTGGCTTTTCTGCTGCTCGAAGATCGGCTTGGAGC 247
Qy 295 AGCGAGCCACCGCTCGGCAAGCCGAGACTCTGTAGGCTTCTCGAATTCGCTGACCC 354
Db 248 AGCGAGCCACCGCTCGGCAAGCCGAGACTCTGTAGGCTTCTCGAATTCGCTGACCC 307
Qy 355 TCAGAGCGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGAGATGAGAGC 414
Db 308 TCAGAGCGCTGAGCGCGCGGCTTACCTGAGAGACTGTCAAGAAAAAGAGAGAGAGC 367
Qy 415 GGGGACAGGCGGATTCGCGGAAACGGCTTGGCCCTCGGGCGGGCTTCCGGTTCTGGCCACC 474
Db 368 GGGGACAGGCGGATTCGCGGAAACGGCTTGGCCCTCGGGCGGGCTTCCGGTTCTGGCCACC 427
Qy 475 CTTTTTCCTCGGAGATCGCAAGCAGGCTCTTCTAGTTTCCGACGCTCTTCCGCGCGGA 534
Db 428 CTTTTTCCTCGGAGATCGCAAGCAGGCTCTTCTAGTTTCCGACGCTCTTCCGCGCGGA 487
Qy 535 AAACTCCGGGAACCCACACTGCTTTCTCTGCGCCAGCCGAGACTCGGGTCACTTACTG 594
Db 488 AAACTCCGGGAACCCACACTGCTTTCTCTGCGCCAGCCGAGACTCGGGTCACTTACTG 547
Qy 595 GAGCAAACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGCTTCCAGAAAGTGTCTAAT 654
Db 548 GAGCAAACTGCTCTCCAGCTCTTGGCGCGCTCCCGGATTCGCTTCCAGAAAGTGTCTAAT 607
Qy 655 TTGGAGCCAACTTTTCGGTGGAAATGTTCCGACCCAGATGGCTAGATTTTCTCGAGTCTA 714
Db 608 TTGGAGCCAACTTTTCGGTGGAAATGTTCCGACCCAGATGGCTAGATTTTCTCGAGTCTA 667
Qy 715 CAGCGCCCTCAGAGCCCTGAAAGGAGGAGAAACAGCCGCCGCCCAACAGCGCAGAAATC 774
Db 668 CAGCGCCCTCAGAGCCCTGAAAGGAGGAGAAACAGCCGCCGCCCAACAGCGCAGAAATC 727
Qy 775 TTTGAGTTGCTGCTGAGCTGACTCTCAGACCCCTCGGTACCAGTCCCTTGAATGGCT 834
Db 728 TTTGAGTTGCTGCTGAGCTGACTCTCAGACCCCTCGGTACCAGTCCCTTGAATGGCT 787
Qy 835 AGAGGAGGGGATCCACTGGCAATACCTGCGCCCCAGACCTTAAATTTGAGCTTAAAGGCCAA 894
Db 788 AGAGGAGGGGATCCACTGGCAATACCTGCGCCCCAGACCTTAAATTTGAGCTTAAAGGCCAA 847
Qy 895 GGGAAAGTGTCTTTGGACCCCTGACACAGAGCTTTTCTCTTAGAGCAGAGCTGT- GGGGAG 953
Db 848 GGGAAAGTGTCTTTGGACCCCTGACACAGAGCTTTTCTCTTAGAGCAGAGCTGT- GGGGAG 907
Qy 954 TGAGCTGTGTCAGTAGCTTCAATCCGCTGTACTCTAAACCGGGAACCTTGGCTCTT 1013
Db 954 TGAGCTGTGTCAGTAGCTTCAATCCGCTGTACTCTAAACCGGGAACCTTGGCTCTT 1013
```

Db 908 TGGAGCTGTGTCAGTAGCTTCAATCMTCTGTCTCTCTTAACWGGGAMTTGGCTCTY 967
QY 1014 CGCCCTCTGGGC 1025
Db 968 GCCCTCTTGGGC 979
RESULT 5
LOCUS BUI62260 897 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7939455 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6172811
5', mRNA sequence.
ACCESSION BUI62260
VERSION BUI62260.1 GI:22676170
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 897)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM13543 row: f column: 12
High quality sequence stop: 673.
Location/Qualifiers
1..897
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6172811"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 29.5%; Score 868.8; DB 5; Length 897;
Best Local Similarity 99.1%; Pred. No. 2.9e-233;
Matches 884; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 21 CGCACCGCGGCTACCCAGTCTTCGCGTATCGCGTGTCTCAGGGGCTTTCAACGCT 80
Db 6 CGTCGAGCGCGCTACCCAGTCTTCGCGTATCGCGTGTCTCAGGGGCTTTCAACGCT 65
QY 81 CTGTCAGTGCGAAACCAATCCGCGAGGCGGTGGGGGACTCCTATCCATGTGTGAAGC 140
Db 66 CTGTCAGTGCGAAACCAATCCGCGAGGCGGTGGGGGACTCCTATCCATGTGTGAAGC 125
QY 141 GTCGAGCGGCTAGGGAACTCTTCCCGCAGGATGGAAGTCGATCAGTCGCGGCT 200
Db 126 GTCGAGCGGCTAGGGAACTCTTCCCGCAGGATGGAAGTCGATCAGTCGCGGCT 185
QY 201 ATTGGCGGGCTGTCTTCCTGTGTCTTCGCGCGGCTGCGGATTCGCTGCCCTCTGT 260
Db 186 ATTGGCGGGCTGTCTTCCTGTGTCTTCGCGCGGCTGCGGATTCGCTGCCCTCTGT 245
QY 261 GGCTTTTCTGTGGCTCGAAGATCGGCTGAGCAGACGCCACCGCTGGCGAAGGCCG 320
Db 246 GGCTTTTCTGTGGCTCGAAGATCGGCTGAGCAGACGCCACCGCTGGCGAAGGCCG 305

QY 321 AGACTCTGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGCCCTA 380
Db 306 AGACTCTGTAGGCTTCTCCGAATCCGTCGACCTCCAGCCGCTGAGCGCGCGCCCTA 365
QY 381 CTTGAGAGACTGTCAAGAAAAAGGAGATGAGCGGGGACAGCGGATTCGGGAAACGGC 440
Db 366 CTTGAGAGACTGTCAAGAAAAAGGAGATGAGCGGGGACAGCGGATTCGGGAAACGGC 425
QY 441 TTGGCCCTCGGGGGGCTTCGGTTCTGGCCACCTTTTCCCTCGCGGATTCGCAAGCAG 500
Db 426 TTGGCCCTCGGGGGGCTTCGGTTCTGGCCACCTTTTCCCTCGCGGATTCGCAAGCAG 485
QY 501 GCTCTTCTAAGTTCCCGACGCTCTTTGGCCCGGAAAACTCCGGGAAACCCACACTGCTTT 560
Db 486 GCTCTTCTAAGTTCCCGACGCTCTTTGGCCCGGAAAACTCCGGGAAACCCACACTGCTTT 545
QY 561 CTTCTGCCAGCCCGAGACTCGGGTCACTTACTGAGCAAACTGCTCTCCAGCTCTCTTG 620
Db 546 CTTCTGCCAGCCCGAGACTCGGGTCACTTACTGAGCAAACTGCTCTCCAGCTCTCTTG 605
QY 621 CGCGCTCCCGGATTCGTTTCAAGAGGTGCTAAATTTGGAGCCAACTTTTCGGTGAATGT 680
Db 606 CGCGCTCCCGGATTCGTTTCAAGAGGTGCTAAATTTGGAGCCAACTTTTCGGTGAATGT 665
QY 681 TTCCGACCATGCTAGATTTTCTGAGTCTACAGCCCTGAGAGCCCTGAAGGAC 740
Db 666 TTCCGACCATGCTAGATTTTCTGAGTCTACAGCCCTGAGAGCCCTGAAGGAC 725
QY 741 GGAGAAACAGCGGCGCCCGCAGCGCAGAAATCTTTGAGTTGCTGAGCTGAGCTCCT 800
Db 726 GGAGAAACAGCGGCGCCCGCAGCGCAGAAATCTTTGAGTTGCTGAGCTGAGCTCCT 785
QY 801 CAGACCCCTCGTCAACAGTCCCTTGAATGGCTAGAGGAGGGATCCACTGGCAATACT 860
Db 786 CAGACCCCTCGTCAACAGTCCCTTGAATGGCTAGAGGAGGGATCCACTGGCAATACT 845
QY 861 CGCCCCCAGACCTAAATTTGGAGCTTAAGCCCAAGGAAAGTGTCTTGGACC 911
Db 846 CGCCCCCAGACCTAAATTTGGAGCTTAAGCCCAAGGAAAGTGTCTTGGACC 897
RESULT 6
LOCUS BQ221596 873 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7569589 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043767
5', mRNA sequence.
ACCESSION BQ221596
VERSION BQ221596.1 GI:20402996
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 873)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM13285 row: e column: 16
High quality sequence stop: 653.
Location/Qualifiers
1..873
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

<div><div>/clone="IMAGE:6043767"</div><div>/tissue_type="embryonal carcinoma, cell line"</div><div>/lab_host="DH10B (phage-resistant)"</div><div>/clone_lib="NIH_MGC 92"</div><div>/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."</div></div>									
ORIGIN	Query Match	28.4%	Score 834.8;	DB 5;	Length 873;				
	Best Local Similarity	99.5%;	Pred. No. 1.2e-223;						
	Matches 858;	Conservative	0;	Mismatches 2;	Indels 2;	Gaps 2;			
Qy	1488	CTGAAGAAAAATAGAAATTAACTACAGAGGTTCCACTTGTCTTTGGAAAGAGAGAGGCC	1547						
Db	1	CTGAGGAAAAATAGAAATTAACTACAGAGGTTCCACTTGTCTTTGGAAAGAGAGAGGCC	60						
Qy	1548	CTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGAAGAGGAGCCTGGAGAGGGCCGAA	1607						
Db	61	CTTCTGAGGCTGTCCATCTAGTGAGATACCTATGGAAGAGGAGCCTGGAGAGGGCCGAA	120						
Qy	1608	TAAGTGTAGTTGATTACTATCCTAGAAAGTGACCTTCCCATTTCTGCCAGAGCAGCTT	1667						
Db	121	TAAGTGTAGTTGATTACTATCCTAGAAAGTGACCTTCCCATTTCTGCCAGAGCAGCTT	180						
Qy	1668	GTAGTAAACAACTGATAGATTATATTTTGGAGGTGCATCCAGTGACCTCGAAACAAGTT	1727						
Db	181	GTAGTAAACAACTGATAGATTATATTTTGGAGGTGCATCCAGTGACCTCGAAACAAGTT	240						
Qy	1728	CTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATGATGGTTTTGTATGATGATA	1787						
Db	241	CTGATCCAGAAGGTGAGGATTGGGATGAGGAAGCTGAGGATGATGGTTTTGTATGATGATA	300						
Qy	1788	GCTCAGTGTGAGACTCAGACCTTGAAACAGACCTGAGGGCTTCACCTTTGGAACCTCTT	1847						
Db	301	GCTCAGTGTGAGACTCAGACCTTGAAACAGACCTGAGGGCTTCACCTTTGGAACCTCTT	360						
Qy	1848	TCTGCACTGTAGATCCCTTATAATCCCCAGAACTTTACAGCAAACTTCAGACTGCTGCCA	1907						
Db	361	TCTGCACTGTAGATCCCTTATAATCCCCAGAACTTTACAGCAAACTTCAGACTGCTGCCA	420						
Qy	1908	GAATTTGTCCTGGAAGACCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAG	1967						
Db	421	GAATTTGTCCTGGAAGACCTTCTGATTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAG	480						
Qy	1968	AGAATTCCTCCAGTCTGGAGCCCTTCTCAGACCCCTGAGCATAGTTCTGGGGAGGAAG	2027						
Db	481	AGAATTCCTCCAGTCTGGAGCCCTTCTCAGACCCCTGAGCATAGTTCTGGGGAGGAAG	540						
Qy	2028	ATGACTGGGAATCTAGTGCAGATGAACGAGAGTCTCAAACTGTGGAACTCATTTCTGTA	2087						
Db	541	ATGACTGGGAATCTAGTGCAGATGAACGAGAGTCTCAAACTGTGGAACTCATTTCTGTA	600						
Qy	2088	ATTCCTGATGACCCCTACAACCCCTTAAATTTTAAAGGCTCCCTTTTCAAACATCAGGGAAA	2147						
Db	601	ATTCCTGATGACCCCTACAACCCCTTAAATTTTAAAGGCTCCCTTTTCAAACATCAGGGAAA	660						
Qy	2148	ATGAGAAAGGCTGTCTGACTCAAAAGACCCCATCTGAGTCGATTTGTGGCCATTTCTGAGT	2207						
Db	661	ATGAGAAAGGCTGTCTGACTCAAAAGACCCCATCTGAGTCGATTTGTGGCCATTTCTGAGT	720						
Qy	2208	GTCAACCTTACTTTTCTTGTAAAGGTGCAGCTGTTTGGGGAGCC - AAGAAAGTGAATGTCCA	2266						
Db	721	GTCAACCTTACTTTTCTTGTAAAGGTGCAGCTGTTTGGGGAGCCAAAGAAAGGGAAATGTCCA	780						
Qy	2267	GACTTCGGTACAGCGTGACGTTCTTTCTTCGGAGGAGACACACATGTCAAAAAGAAAAAG	2326						
Db	781	GACTTCGGTACAGCGTGACGTTCTTTCTTCGGAGGAGACACACCCATGTCAAAAGAAAAAG	840						
Qy	2327	GTAA - CCTTCTCTGGAAGGTT 2347							

Db 481 AGGTGCATCCAGTGACCTGGAAACAAGTTCTGATCCAGAGGTGAGGATTTGGGATGAGGA 540
 Qy 1759 AGCTGAGGATGATGGTTTGTAGTAGTAGTCACTCTCAGACTCAGACTCTGGAACAAGA 1818
 Db 541 AGCTGAGGATGATGGTTTGTAGTAGTAGTCACTCTCAGACTCAGACTCTGGAACAAGA 600
 Qy 1819 CCCTGAAGGGCTTCACTTTGGAACTCTTTCTGAGTGTGATGCTTTATATATCCCAAGAA 1878
 Db 601 CCCTGAAGGGCTTCACTTTGGAACTCTTTCTGAGTGTGATGCTTTATATATCCCAAGAA 660
 Qy 1879 CTTTACAGCAACAACTCAGACTGCTGCCAGAAATGTTCTCTG-AGAGCCCTTCTGATTCAG 1937
 Db 661 CTTTACAGCAACAACTCAGACTGCTGCCAGAAATGTTCTCTGAAAGAGCCCTTCTGATTCAG 720
 Qy 1938 AGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATCTCTCCAGCTCTGGAAGCCCTTCTG 1997
 Db 721 AGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAATCTCTCCAGTCTGGAGCCCTTCCNN 780
 Qy 1998 AGA---CCCTGAGCAGTAGTT-CTGGGGAGGAAGA--TGACTGGGAATCTAGTGCAGATG 2051
 Db 781 TGAGACCCCTCAGCAGTAGTTCTCTGGGGAGGAAGAATGACTGGNGAATCTAGTGCAGATG 840
 Qy 2052 AAGCAG-AGAGTCTCAACTGTGGAATCTATCTGTAAATCTGTATGA-----CCCTTACA 2105
 Db 841 AAGCANAAGAGTCTCAACTGTGGAATCTATCTGTAAATCTGGATGACCCCTTAACA 900
 Qy 2106 ACCCTTTAAA---TTTTAAGGCTCTTTTCAACATCAGGGGAAATGAGAAA 2155
 Db 901 ACCCTTTAAAATTTTAAAGGCTCTTTTTCNACTTCAGGGGGAAGAAA 953

RESULT 9
 CD642687
 LOCUS
 DEFINITION AGENCOURT_14569880 NIA Human H1 Embryonic Stem Cell cDNA Library (Long) Homo sapiens cDNA clone IMAGE:30427369 5', mRNA sequence.
 ACCSSION CD642687
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 775)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: csapbe@mail.nih.gov
 Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
 cDNA Library Preparation: Yulan Piao and Minoru Ko
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC c lone distribution information
 can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM514 row: 1 column: 02
 High quality sequence stop: 730.
 FEATURES
 Location/Qualifiers
 1..775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30427369"
 /tissue_type="Embryonic Stem cells"
 /cell_line="WA01"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"
 /note="Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI;

ORIGIN

Query Match 26.1%; Score 768.2; DB 6; Length 775;
 Best Local Similarity 99.4%; Pred. No. 7.1e-205;
 Matches 770; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 356 CAGCGCGCTGAGCGCGCGCCCTACTCAGAGACTGTCAAGAAAAGAGATGGAGCCG 415
 Db 1 CAGCGCGCTGAGCGCGCGCCCTACTCAGAGACTGTCAAGAAAAGAGATGGAGCCG 60
 Qy 416 GGCACAGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGCGCTTCGGCTTCGCCACCC 475
 Db 61 GGCACAGGCGGATCGCGGAAACGGCTTGGCCCTCGGCGCGCTTCGGCTTCGCCACCC 120
 Qy 476 TTTTTCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCCAGCGCTCTTGGCCCGAA 535
 Db 121 TTTTTCCTCGGCGATCGCAAGCAGGCTCTTCTAAGTTCGCCAGCGCTCTTGGCCCGAA 180
 Qy 536 AACTCGGGAAACCCACACATGCTTCTCTGCCAGCGCGAGACTCGGTCAGTACTGG 595
 Db 181 AACTCGGGAAACCCACACATGCTTCTCTGCCAGCGCGAGACTCGGTCAGTACTGG 240
 Qy 596 ACCAATCTCTCTCCAGCTCTTGGCGCGCTCCCGGATTCCTCAGAGAGGTGCTAATT 655
 Db 241 ACCAATCTCTCTCCAGCTCTTGGCGCGCTCCCGGATTCCTCAGAGAGGTGCTAATT 300
 Qy 656 TGGAGCAACATTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTCTGGAGTCTAC 715
 Db 301 TGGAGCAACATTTTCGGTGGAAATGTTTCCGACCAAGATGGCTAGATTTTCTGGAGTCTAC 360
 Qy 716 AGCGCCTCAGAGCCCTGAAAGGAGCGGAGAAAACAGCGCGCCGCCACAGCGCAAAATCT 775
 Db 361 AGCGCCTCAGAGCCCTGAAAGGAGCGGAGAAAACAGCGCGCCGCCACAGCGCAAAATCT 420
 Qy 776 TTGAGTTCCGCTGCAGCTCGACTCTCAGACCCCTCGGTACACAGTCCCTTGTGTTGGCTA 835
 Db 421 TTGAGTTCCGCTGCAGCTCGACTCTCAGACCCCTCGGTACACAGTCCCTTGTGTTGGCTA 480
 Qy 836 GAGGAGGGGATCCACTGCGCAATACTCGCCGCCACAGACTTAAATTTGGAGCTTTAAGGCCAAG 895
 Db 481 GAGGAGGGGATCCACTGCGCAATACTCGCCGCCACAGACTTAAATTTGGAGCTTTAAGGCCAAG 540
 Qy 896 GGAAGTGTCTTGGACCCCTGCAGCAGAGGCTTTTCTCTTAGAGCAGCAGCTGTGGGGAGTG 955

This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199] from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on MBF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTP, TERT, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1, TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days after plating), the ES cells from 4 X 6cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID: 11544199] Double-stranded cDNAs were synthesized with an Oligo(dT) primer [invitrogen: 5'-pGCTAGTCTAGATCGAGCGCGCCCTTTTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb.

```
Db 541 GGAAGTGTCTTGGAGCCCTGCAGCAGAGGCTTTCTCTTAGAGCAGCAGCTGTGGGAGTG 600
Qy 956 GAGCTGTTGGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCCGGAACTTGGCTCTTCG 1015
Db 601 GAGCTGTTGGCCAGTAGCTTCAATCCCGTCTGTACTCTTAACCCGGAACTTGGCTCTTCG 660
Qy 1016 CCCTCTGGGCGCTTAAACATTCACCGCATAGACAATTTTCAGTGTGGTATCCTATTTCGTG 1075
Db 661 CCCTCTGGGCGCTTAAACATTCACCGCATAGACAATTTTCAGTGTGGTATCCTATTTCGTG 720
Qy 1076 AACCTTCTCTACCTAGGACTGCTTCTTCTAGGCTAGAGTACAGTATACGAAACAGTG 1130
Db 721 AACCTTCTCTACCTAGGACTGCTTCTTCTTANGCTAGAACTAGAGTACAGTATACGAAACAGTG 775
```

```
RESULT 10
CA749002/c
LOCUS
DEFINITION
  UI-H-FEI-bei-a-04-0-UI.s2 NCI CGAP FE1 Homo sapiens cDNA clone
  UI-H-FEI-bei-a-04-0-UI 3', mRNA sequence.
ACCESSION
  CA749002
VERSION
  CA749002.1 GI:25569930
KEYWORDS
  EST.
ORGANISM
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 793)
REFERENCE
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: James Martin
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bento-soares@iowa.edu
  The following repetitive elements were found in this cDNA
  sequence: 1-22, >AT rich#Low_complexity
  Seq primer: M13 FORWARD
  POLYA=Yes.
```

```
FEATURES
  source
    1..793
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-H-FEI-bei-a-04-0-UI"
    /tissue_type="Cell lines"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="NCI CGAP_FEI"
    /note="Organ: Chondrosarcoma; Vector: p77T3-Pac
    (Pharmacia) with a modified polylinker; Site 1: Ecor I;
    Site 2: Not I; NCI CGAP FE1 is a normalized cDNA library
    derived from a pool of mRNA obtained from 3 cell lines
    from grade II chondrosarcoma tissues. The library was
    constructed according to Bonaldo, Lennon and Soares,
    Genome Research, 6:791-806, 1996. First strand cDNA
    synthesis was primed with an oligo-dT primer containing a
    Not I site. Double stranded cDNA was ligated to an Ecor I
    adaptor, digested with Not I, and cloned directionally
    into p77T3-Pac vector. The oligonucleotide used to prime
    the synthesis of first-strand cDNA contains a library tag
    sequence that is located between the Not I site and the
    (dT)18 tail. The sequence tag for this library is
    CGCTACGGAC. The cell lines were provided by Dr James
    Martin from the University of Iowa.
    TAG TISSUE=Human grade 2 chondrosarcoma cell line pool
    TAG_L1B=UI-H-FEI
```

```
ORIGIN
  TAG_SEQ=CGCTACGGAC"
  Query Match 25.88; Score 758; DB 6; Length 793;
  Best Local Similarity 99.1%; Pred. No. 5.5e-202;
  Matches 761; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1570 TGAGATACCTATGCGAAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTTGATTACTCAT 1629
Db 793 TGAGATACNTATGCGAAAGAGAGCCTGGAGAGGGCCGAATAAGTGTAGTTGATTACTCAT 734
Qy 1630 CCTAGAAAGGTGACCTTCCCAATTTCTGCCAGACACAGCTTGTAGTAAACAACTGATAGATTA 1689
Db 733 CNTAGAAAGGTGACCTTCCCAATTTCTGCCAGACACAGCTTGTAGTAAACAACTGATAGATTA 674
Qy 1690 TATTTTGGGAGGTGCATCCAGTGACCTGGGAAACAAAGTTCTGATCCACAAGGTGAGGATTTG 1749
Db 673 TATTTTGGGAGGTGCATCCAGTGACCTGGGAAACAAAGTTCTGATCCACAAGGTGAGGATTTG 614
Qy 1750 GGATGAGGAAGCTCAGGATGATGGTTTTGTAGTAGTGTAGTCTCACTGTCCAGACTCAGACCT 1809
Db 613 GGATGAGGAAGCTCAGGATGATGGTTTTGTAGTAGTGTAGTCTCACTGTCCAGACTCAGACCT 554
Qy 1810 TGAACAAAGACCTGGAAGGGCTTCACTTTGGAACTCTTTCTGCAGTGTAGATCCCTTATAA 1869
Db 553 TGAACAAAGACCTGGAAGGGCTTCACTTTGGAACTCTTTCTGCAGTGTAGATCCCTTATAA 494
Qy 1870 TCCCCAGAACTTTTACAGCAAAATTCAGACTGCTGCCAGAAATTTCTCTGAAGAGCCTTC 1929
Db 493 TCCCCAGAACTTTTACAGCAAAATTCAGACTGCTGCCAGAAATTTCTCTGAAGAGCCTTC 434
Qy 1930 TGATTTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAG 1989
Db 433 TGATTTCAGAGAAGGATTTGTCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAG 374
Qy 1990 CCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTGCAGA 2049
Db 373 CCTTCTGAGACCCCTGAGCATAGTTCTGGGGAGGAAGATGACTGGGAATCTAGTGCAGA 314
Qy 2050 TGAAGCAGAGAGTCTCAAACTGTGGAACTCATTTCTGTAAATTCGATGACCCCTACAAACCC 2109
Db 313 TGAAGCAGAGAGTCTCAAACTGTGGAACTCATTTCTGTAAATTCGATGACCCCTACAAACCC 254
Qy 2110 TTTAAATTTTAAAGGCTCCTTTTCAACATCAGGGGAAAAATGAGAAAGGCTGTGCTGATC 2169
Db 253 TTTAAATTTTAAAGGCTCCTTTTCAACATCAGGGGAAAAATGAGAAAGGCTGTGCTGATC 194
Qy 2170 AAAGACCCCATCTGAGTCCATTTGGCCATTTCTGAGTGTCCACACCTTACTTTCTGTAA 2229
Db 193 AAAGACCCCATCTGAGTCCATTTGGCCATTTCTGAGTGTCCACACCTTACTTTCTGTAA 134
Qy 2230 GGTGCAGCTGTGGGGAGGCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTTCT 2289
Db 133 GGTGCAGCTGTGGGGAGGCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTTCT 74
Qy 2290 TTCTGGAGGAAGACACACATGTCAAAAGAAAAAGGTAACTTCCT 2337
Db 73 TTCTGGAGGAAGACACACATGTCAAAAGAAAAAGGTGTGTTCTT 26
RESULT 11
CA426770/c
LOCUS
DEFINITION
  UI-H-FEI-bei-a-04-0-UI.s1 NCI CGAP FE1 Homo sapiens cDNA clone
  UI-H-FEI-bei-a-04-0-UI 3', mRNA sequence.
ACCESSION
  CA426770
VERSION
  CA426770.1 GI:24789496
KEYWORDS
  EST.
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 795)
REFERENCE
```

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-22, >AT rich#Low_complexity
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source Location/Qualifiers
1. .795
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FEI-bel-a-04-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FE1"
/notes="Organ: Chondrosarcoma; Vector: p7T3-Pac
(Pharmacia) with a modified polylinker; Site 1: EcoR I;
Site 2: Not I; NCI CGAP FE1 is a normalized cDNA library
derived from a pool of mRNA obtained from 3 cell lines
from grade II chondrosarcoma tissues. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into p7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCTACGGAC. The cell lines were provided by Dr James
Martin from the University of Iowa.
TAG TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG LIB=UI-H-FEI
TAG_SEQ=CGCTACGGAC"

ORIGIN
Query Match 25.7%; Score 755.6; DB 6; Length 795;
Best Local Similarity 98.8%; Pred. No. 2.6e-201;
Matches 761; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1568 AGTGAGATACCTATGGAAGAGGAGCCCTGGAGAGGCGCGAATAAGTGTAGTTGATTACTCA 1627
DB 795 AGTGAGATACCTATGGAAGAGGAGCCCTGGAGAGGCGCGAATAAGTGTAGTTGATTACTCA 736
QY 1628 TACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCACTGTGTAGTAAACAACTGATAGAT 1687
DB 735 TACCTAGAAGGTGACCTTCCCATTTCTGCCAGACCACTGTGTAGTAAACAACTGATAGAT 676
QY 1688 TATATTTGGAGGTGATCCAGTGACCTGGAACAAAGTTCTGTATCCAGAGGTGAGGAT 1747
DB 675 TATATTTGGAGGTGATCCAGTGACCTGGAACAAAGTTCTGTATCCAGAGGTGAGGAT 616
QY 1748 TGGATCAGGAAGCTGAGGATGATGGTTTGTATAGTATAGCTCAGTGTGAGCTCAGAC 1807
DB 615 TGGATCAGGAAGCTGAGGATGATGGTTTGTATAGTATAGCTCAGTGTGAGCTCAGAC 556
QY 1808 CTTGAACAAGACCTGAAGGCTTTCACCTTTGGAACTCTTTCTGACGTAGATCCTTAT 1867
DB 555 CTTGAACAAGACCTGAAGGCTTTCACCTTTGGAACTCTTTCTGACGTAGATCCTTAT 496
QY 1868 AATCCCAAGAACTTTACAGCAACAAATTCAGACTCTGCGAGATTTGTTCTCGAAGCCT 1927

Db 495 AATCCCCAGAACTTTACAGCAACAATTACAGACTGCTGCCAGAAATTTGTTCTGGAAGCCT 436
QY 1928 TCTGATTTCAGAGAAGGATTTGCTGCAAGTCTGTATCTAGAGAAATTCCTCCAGTCTGGA 1987
DB 435 TCTGATTTCAGAGAAGGATTTGCTGCAAGTCTGTATCTAGAGAAATTCCTCCAGTCTGGA 376
QY 1988 AGCTTCTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGAATCTAGTGCA 2047
DB 375 AGCTTCTCTGAGACCCCTGAGCATAGTTCTGGGAGGAAGATGACTGGGAATCTAGTGCA 316
QY 2048 GATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATCTGTAAATTTCTGATGACCCCTACAAC 2107
DB 315 GATGAAGCAGAGAGTCTCAAACTGTGGAACCTCATCTGTAAATTTCTGATGACCCCTACAAC 256
QY 2108 CTTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAAAATGAGAAAGCTGTCTGTGAC 2167
DB 255 CTTTAAATTTTAAAGGCTCTCTTTTCAACATCAGGGGAAAAATGAGAAAGCTGTCTGTGAC 196
QY 2168 TCAAGACCCCATCTGAGTCCATTTGCGGCATTTCTGAGTGTACACCTTACTTTCTTGT 2227
DB 195 TCAAGACCCCATCTGAGTCCATTTGCGGCATTTGAGTGTACACCTTACTTTCTTGT 136
QY 2228 AGGTGCAGCTGTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTT 2287
DB 135 AGGTGCAGCTGTGGGAGCCCAAGAAAGTGAATGTCCAGACTCGGTACAGCGTGACGTT 76
QY 2288 CTTTCTGAGGAGACACACATGTCATAAAGAAAAAGGTAACTTCTCT 2337
DB 75 CTTTCTGAGGAGACACACATGTCATAAAGAAAAAGGTAACTTCTCT 26

RESULT 12
BG721970/c
LOCUS BG721970 850 bp mRNA linear EST 08-MAY-2001
DEFINITION G02698720F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4830749 5',
mRNA sequence.
ACCESSION BG721970
VERSION BG721970.1 GI:14001157
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 850)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10752 row: g column: 06
High quality sequence stop: 809.
Location/Qualifiers
1. .850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4830749"
/lab_host="DH10B"
/clone_lib="NIH_MGC_97"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and

FEATURES
source
1. .850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4830749"
/lab_host="DH10B"
/clone_lib="NIH_MGC_97"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and

constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN	Query Match	Best Local Similarity	Score	748.8;	DB 4;	Length	850;
	Matches	811;	Conservative	0;	Mismatches	12;	Indels
Qy	655	TTGGAGCCAACTTTTCGGTGGAAATGTTTCCGACAG-ATGCTAGA-TTTTGTCTGGAGTC	712				
Db	832	TTGGACCAACTTTTCGGTGGAAATGTTTCCGACAGATGGCTAGATTTTTCGTGGATTC	773				
Qy	713	T-ACAGCCCTGAGAGCCCTGAGGAGCGGA-GAACACGACCGCCGCCACACGGCAGA	770				
Db	772	TAAACACCCCTGAGACCCTGAAAGGAGCGGAGAAACCAACGCCGCCACACGGCAGA	713				
Qy	771	AATCTTTGAGTTCGTCGAGCTCGACTCCTCAGACCCCTCGGTCAACAGTCCCTTGATT	830				
Db	712	AATCTTTGAGTTCGTCGAGCTCGACTCCTCAGACCCCTCGGTCAACAGT-CCCTTGATT	654				
Qy	831	GGCTAGAGGAGGATCCACTGGCAATCTCGCCCCCAGACTTAAATTTGAGCTTAAGG	890				
Db	653	GGCTAGAGGAGGATCCACTGGCAATCTCGCCCCCAGACTTAAATTTGAGCTTAAGG	594				
Qy	891	CCAAGGGAAGTGCTTTGGACCTCGCAGCACAGGCTTTTCTTACGAGCAGCAGCTGTGG	950				
Db	593	CCAAGGGAAGTGCTTTGGACCTCGCAGCACAGGCTTTTCTTACGAGCAGCAGCTGTGG	534				
Qy	951	GAGTGGAGCTTTGCCAGTAGTCTTCAATCCCGTCTGTAATCTTAAACCGGAACTTGCT	1010				
Db	533	GAGTGGAGCTTTGCCAGTAGTCTTCAATCCCGTCTGTAATCTTAAACCGGAACTTGCT	474				
Qy	1011	CTTCGCCCTCTGGGCTCTTAAACATTTCAACGCATAGACAATTTCAAGTGTGTATTCCTATT	1070				
Db	473	CTTCGCCCTCTGGGCTCTTAAACATTTCAACGCATAGACAATTTCAAGTGTGTATTCCTATT	414				
Qy	1071	TGCTGMAACCTTCTTACCTGACCTGCTTTCTTAGCTAGAGTCAAGTATCAGAACAGTG	1130				
Db	413	TGCTGMAACCTTCTTACCTGACCTGCTTTCTTAGCTAGAGTCAAGTATCAGAACAGTG	354				
Qy	1131	ATGGAATPAGGAGTAGTTCGGCTTCCAGACACTAACCACAGAGAGCAGCTGCCTGAGAG	1190				
Db	353	ATGGAATPAGGAGTAGTTCGGCTTCCAGACACTAACCACAGAGAGCAGCTGCCTGAGAG	294				
Qy	1191	AGGACCATTTGATCCCGAGCCGCTGAGTCAGAACTCATTTCCGGCCCTCTGGCAGGAT	1250				
Db	293	AGGACCATTTGATCCCGAGCCGCTGAGTCAGAACTCATTTCCGGCCCTCTGGCAGGAT	234				
Qy	1251	GTCCACCTCTTTTACGGAAGGCTTACAGAAATTTTCAACATCTTCGCATGAAACGGCTGG	1310				
Db	233	GTCCACCTCTTTTACGGAAGGCTTACAGAAATTTTCAACATCTTCGCATGAAACGGCTGG	174				
Qy	1311	AATTCCTTTCAACAGGCTTAAAGGGGCAAGATTTTACCCACCCCTGACCAAGATTAATGGCT	1370				
Db	173	AATTCCTTTCAACAGGCTTAAAGGGGCAAGATTTTACCCACCCCTGACCAAGATTAATGGCT	114				
Qy	1371	ACCAAGCTGGAGGAGAACACAGCCTTCTCCGATGGATGCAAAACACTGACAGATA	1430				
Db	113	ACCAAGCTGGAGGAGAACACAGCCTTCTCCGATGGATGCAAAACACTGACAGATA	54				
Qy	1431	ACCAACACAGTTTGTCTCTGCTGGAGACATTTCTGGAAACACCC	1478				
Db	53	ACCAACACAGTTTGTCTCTGCTGGAGACATTTCTGGAAACGCC	6				

RESULT 13
 BQ427697
 LOCUS
 DEFINITION AGNCOURT_7898324 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6154236
 5', mRNA sequence.
 ACCESSION BQ427697

BQ427697.1 GI:21166773
 EST.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 875)
 NTH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: rgs@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M13494 row: p column: 13
 High quality sequence stop: 557.
 Location/Qualifiers
 1..875
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6154236"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_67"
 /notes="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Library constructed by Life Technologies."

ORIGIN

Query Match	25.0%;	Score	736.4;	DB 5;	Length	875;
Best Local Similarity	96.2%;	Pred. No.	7.2e-196;			
Matches	776;	Conservative	0;	Mismatches	27;	Indels
Qy	1708	CAGTACCTGGAAACAAGTTCTGATCCAGAAGTGGAGATTGGATGAGGAGCTGAGGA	1767			
Db	1	CAGTACCTGGAAACAAGTTCTGATCCAGAAGTGGAGATTGGATGAGGAGCTGAGGA	60			
Qy	1768	TGATGTTTGTAGTAGTAGTCACTGTGAGACTCAGACTTGAACAAGACCTGGAAGG	1827			
Db	61	TGATGTTTGTAGTAGTAGTCACTGTGAGACTCAGACTTGAACAAGACCTGGAAGG	120			
Qy	1828	GCTTACCTTTGGAACTCTTTCTGCAAGTGTAGATCCCTTATAATCCCAAGACCTTACAGC	1887			
Db	121	GCTTACCTTTGGAACTCTTTCTGCAAGTGTAGATCCCTTATAATCCCAAGACCTTACAGC	180			
Qy	1888	AACAATTCAGACTGCTGCCAGAAATTTCTCTGAGAGCCCTTCTGATTCAGAGGAGTTT	1947			
Db	181	AACAATTCAGACTGCTGCCAGAAATTTCTCTGAGAGCCCTTCTGATTCAGAGGAGTTT	240			
Qy	1948	GTCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGA	2007			
Db	241	GTCTGGCAAGTCTGATCTAGAGAAATTCCTCCAGTCTGGAAGCCTTCTGAGACCCCTGA	300			
Qy	2008	GCATAGTTCTGGGAGGAAGATCACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAA	2067			
Db	301	GCATAGTTCTGGGAGGAAGATCACTGGGAATCTAGTGCAGATGAAGCAGAGAGTCTCAA	360			
Qy	2068	ACTGTGGAACCTATTCTGTAATTTCTGATGACCCCTACACCCCTTAAATTTAAGGCTCC	2127			
Db	361	ACTGTGGAACCTATTCTGTAATTTCTGATGACCCCTACACCCCTTAAATTTAAGGCTCC	420			
Qy	2128	TTTTTCAACATCAGGGGAAATGAGAAAGGCTGCTGCTGACTCAAAAGACCCCATCTGAGTC	2187			
Db	421	TTTTTCAACATCAGGGGAAATGAGAAAGGCTGCTGCTGACTCAAGAGACCCCATCTGAGTC	480			
Qy	2188	CATTGTGGCCATTCTTCTGAGTGTACACCTTATCTTTCTTTGTAAGGTGACGCTGTGTGGGAG	2247			

```

481 CATTGTGCCATTCTTGAGTGTACACCTTACTTCTTGTAAAGTGCAGCTGTGGGGAG 540
2248 CCAGAAGTGAATGTCAGACTCGGTACAGCGTACGAGTCTTCTTCTCGAGGAGACACAC 2307
541 CCAAGAAGTGAATGTCAGACTCGGTACAGCGTACGAGTCTTCTTCTCGAGGAGACACAC 600
2308 ACATGTCAAAAGAAAAAGGTAACCTTCTTCTTGAAGAGTGTACTGAGTATTATATAAGTGG 2367
601 ACATGTCAAAAGAAAAAGGTAACCTTCTTCTTGAAGAGTGTACTGAGTATTATATAAGTGG 660
2368 TGATGAGATCGAAAGACCATCGGAGAGATTTGCAAGGATGATGAGTTCAGAA 2427
661 TGATGAGATCGCAAGAGACCATCGGAGAGATTTGCAAGGATGATGAGTTCAGAA 720
2428 ACGAATTTCAAGAAACAGAGATGCTATTGG--ATATTGCTTGACATTTGAACACAGAGAA 2485
721 AACAAATTCAGAAACAGAGATGCTATTGGATATTGCTTTGACATTTTGGACCCGAGA 780
2486 AG--AATGTTTAATAGACTCCAGGGAA 2510
781 AGGAATGTTTATAAGACTCCAGGGGA 807

```

RESULT 14
CF594071
LOCUS
DEFINITION AGENCOURT_15624090 NIH_MGC_147 Homo sapiens cDNA clone
IMAGE:30528106 5', mRNA sequence.

ACCESSION CF594071
VERSION CF594071.1 GI:36348197
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 838)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM613 row: i column: 11
High quality sequence stop: 595.
Location/Qualifiers
1. 838
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30528106"
/tissue_type="Human Placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_147"
/notes="Organ: placenta; Vector: pBluescriptR; Site_1:
ali-XhoI; Site_2: BamHI; Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.3 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: This is
a NIH_MGC library."

FEATURES
source

ORIGIN

```

Query Match 25.0%; Score 734.6; DB 7; Length 838;
Best Local Similarity 97.8%; Pred. No. 2.3e-195;
Matches 787; Conservative 0; Mismatches 14; Indels 4; Gaps 4;

QY 67 GGCTTTTCAACCCCTCTGTAGTCGGAAAAACCATCGCGAGGCGCTGGGGGAGACTCTTATC 126
DB 28 GGCTTTTCAACCCCTCTGTAGTCGGAAAAACCATCGCGAGGCGCTGGGGGAGACTCTTATC 87
QY 127 CATGTGTGTTGAAGCGTCGAGCCGACCTAGGGAACCTCTTCCCGCCAGAGATGGAAGTCGC 186
DB 88 CATGTGTGTTGAAGCGTCGAGCCGACCTAGGGAACCTCTTCCCGCCAGAGATGGAAGTCGC 147
QY 187 ATCAGTCGCGCCCTATTGGCGCGGCTGTCTTCCCTGTGTCTTCCCGCCGCTGCCGCAT 246
DB 148 ATCAGTCGCGCCCTATTGGCGCGGCTGTCTTCCCTGTGTCTTCCCGCCGCTGCCGCAT 207
QY 247 TCCTGCTCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCCTGAGAGCAGCAGCCACC 306
DB 208 TCCTGCTCTCTGTGGCTTTTCTGTGGCTCGAAGATCGGCCTGAGAGCAGCAGCCACC 267
QY 307 GCTGGGCNAGGCCGAGACTCTGTAGGCTTCTCCGAATCCCTCGACCTCCAGCCGCTGA 366
DB 268 GCTGGGCNAGGCCGAGACTCTGTAGGCTTCTCCGAATCCCTCGACCTCCAGCCGCTGA 327
QY 367 GCCTCCGCGGCCCTTACCTGAGAGACTGTCAAGAAAAAGGAGATGAGCCGGGACAGCGG 426
DB 328 GCCTCCGCGGCCCTTACCTGAGAGACTGTCAAGAAAAAGGAGATGAGCCGGGACAGCGG 387
QY 427 ATCGCGAAAAAGCGCTTTGGCCCTCGGCGGCTTCCGGTTCTGCGCCACCTTTTCCCTCG 486
DB 388 ATCGCGAAAAAGCGCTTTGGCCCTCGGCGGCTTCCGGTTCTGCGCCACCTTTTCCCTCG 447
QY 487 GCATCGCAGCAGCGCTTCTTAAGTTCCGAGCGCTCTTGGCCCGGAAAACTCCGGGAA 546
DB 448 GCATCGCAGCAGCGCTTCTTAAGTTCCGAGCGCTCTTGGCCCGGAAAACTCCGGGAA 507
QY 547 CCCACACTGCTTTCTCTGCCCGAGAGACTCGGGTCAGTTACTTGAGAGAACTGCT 606
DB 508 CCCACACTGCTTTCTCTGCCCGAGAGACTCGGGTCAGTTACTTGAGAGAACTGCT 567
QY 607 CTCCCAGCTCTTTGCGCGCTCTCCCGGATTTGCTTCAGAGAGTCTTAATTTGGAGCCAACT 666
DB 568 CTCCCAGCTCTTTGCGCGCTCTCCCGGATTTGCTTCAGAGAGTCTTAATTTGGAGCCAACT 627
QY 667 TTTTGGTGGGATTTTTCGACAGATGGCTAGATTTTGTGGAGTCTTACAGCCCTTGAG 726
DB 628 TTTTGGTGGGATTTTTCGACAGATGGCTAGATTTTGTGGAGTCTTACAGCCCTTGAG 687
QY 727 AGCCCTGAAGGGACGGAGAAACAGCCGCCGCCACAGCGCAGAAAATCTTTTGAG-TTCGC 785
DB 688 AGCCCTGAAGGGACGGAGAAACAGCCGCCGCCACAGCGCAGAAAATCTTTTGAGTTCCG 747
QY 786 TGAGTCTGACTCTCTCAGACCCCTC-GGTCCACAGTCCCTTG-ATTGGCTAGAGAGGG 843
DB 748 TGAGTCTGACTCTCTCAGACCCCTC-GGTCCACAGTCCCTTG-ATTGGCTAGAGAGGG 807
QY 844 G-ATCCACTGGCAATACTCGCCCCC 867
DB 808 GAATCCACTGGCAATACTCGCCCCC 832

```

RESULT 15

CB989883
LOCUS
DEFINITION AGENCOURT_13903499 NIH_MGC_147 Homo sapiens cDNA clone
IMAGE:30341796 5', mRNA sequence.
ACCESSION CB989883
VERSION CB989883.1 GI:30284403
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 793)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

JOURNAL

Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM373 row: n column: 13
High quality sequence stop: 654.

FEATURES

source

1..793
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30341796"
/tissue_type="Human Placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH MGC 147"
/note="Organ: placenta; Vector: pbluescriptR; Site_1:
ali-KhoI; Site_2: BamH; Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.3 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein,
(NIH/NHGRI, National Institutes of Health). Note: This is
a NIH_MGC library."

ORIGIN

Query Match 24.8%; Score 728.4; DB 6; Length 793;
Best Local Similarity 97.9%; Pred. No. 1.3e-193;
Matches 738; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCTCTTCGCGTATCGCGTTG 60
Db 28 ATTTTGGGCTTCGCTTCCACCGCACCGCGGCTTACCCAGTCTCTTCGCGTATCGCGTTG 87
QY 61 CTCAGGGGCTTTTCAACCTCTGTGAGTGGGAAACCATCGCGAGGCGGTGGGGGACT 120
Db 88 CTCAGGGGCTTTTCAACCTCTGTGAGTGGGAAACCATCGCGAGGCGGTGGGGGACT 147
QY 121 CCTATCCATGTTGTTGAAGCGTTCGAGCGGACTAGGGAACCTCTTCCCGCCAGGATGGA 180
Db 148 CCTATCCATGTTGTTGAAGCGTTCGAGCGGACTAGGGAACCTCTTCCCGCCAGGATGGA 207
QY 181 AGTCGCATCAGTCGCGGCTATTGCGCGGCTGTTCTTCCCTGTGTTCTGCGCCCGCTG 240
Db 208 AGTCGCATCAGTCGCGGCTATTGCGCGGCTGTTCTTCCCTGTGTTCTGCGCCCGCTG 267
QY 241 CGGATTCGCTGCCCTCTGTGGCTTTTCTGTGGTTCGAAGATCGGCGCTGAGCAGCGAC 300
Db 268 CGGATTCGCTGCCCTCTGTGGCTTTTCTGTGGTTCGAAGATCGGCGCTGAGCAGCGAC 327
QY 301 GCCACGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCCAGC 360
Db 328 GCCACGCTGGGCAAGGCGGAGACTCTGTAGGCTTCTCCGAATCCCGTCGACCTCCAGC 387
QY 361 CGCTGAGCGCGCGGCGCTTACTGAGAGACTGTCAAGAAAAAGGAGATGGAGCCGGGAC 420
Db 388 CGCTGAGCGCGCGGCGCTTACTGAGAGACTGTCAAGAAAAAGGAGATGGAGCCGGGAC 447
QY 421 AGGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGCTTCCGGTCTTGGCCACCTTTT 480
Db 448 AGGCGGATCGCGGAAACGGCTTGGCCCTCGGGCGGCTTCCGGTCTTGGCCACCTTTT 507

QY 481 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTGGCCCGGAAAACTC 540
Db 508 CCCTCGCGGATCGCAAGCAGGCTCTTCTAAGTTCCCGACGCTCTTGGCCCGGAAAACTC 567
QY 541 CGGGAACCCACACTGCTTTTCTTCTGCGCCGAGACTCGGGTCAAGTTACTGAGCGAA 600
Db 568 CGGGAACCCACACTGCTTTTCTTCTGCGCCGAGACTCGGGTCAAGTTACTGAGCGAA 627
QY 601 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCCTTCAGNAGGTGCTAATTTGGAG 660
Db 628 ACTGCTCTCCAGCTCCTTGGCGGCTCCCGGATTCCTTCAGNAGGTGCTAATTTGGAG 687
QY 661 CCAACTTTTTCGGTGAATGTTTCCGACACAGATGGCTAGATTTTCTGAGTCTACAGCGC 720
Db 688 CCAACTTTTTCGGTGAATGTTTCCGACACAGATGGCTAGATTTTCTGAGTCTACAGCGC 747
QY 721 CCTGAGAGCCCTGAAGGACGGGAGAAACGAGCC 754
Db 748 CCTGAGAGCCCTGAAGGACGGGAGAAACGAGCC 781

Search completed: September 16, 2005, 08:56:04
Job time : 8799.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2005, 16:13:49 ; Search time 70.2388 Seconds
(without alignments)
3926.035 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MPEGTGSRKRLGPRAGFRP.....RMFNRLQGTCFKGLNLVKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3804	100.0	713	8	AQ88458 Human GAD
2	3794	99.7	713	4	AAB95876 Human pro
3	3627	95.3	707	3	AAB53401 Human col
4	2223	58.4	698	8	AQ88460 Mouse GAD
5	813	21.4	153	4	AM25448 Human pro
6	260.5	6.8	674	3	AA184366 Amino aci
7	260.5	6.8	674	7	ADE54748 Human pro
8	260.5	6.8	674	7	ADD45119 Human pro
9	260.5	6.8	674	7	ADE54744 Human pro
10	259.5	6.8	674	4	AAB92888 Human pro
11	259.5	6.8	674	8	ADR14089 Human pro
12	232.5	6.6	674	2	AAW98991 Human gro
13	238	6.3	657	2	AA141102 Rat Myd11
14	237	6.2	657	5	ABBS5733 Mouse isc
15	219	5.8	578	2	AAW79958 Human pro
16	219	5.8	578	2	AAW06514 Human pro
17	219	5.8	578	2	AA141104 Human pro
18	201	5.3	590	2	AA141101 Rat gadd3
19	170	4.5	995	8	ADR98899 Lung spec
20	168	4.4	1454	8	ABM84777 Human dia
21	167	4.4	1596	6	ABG73372 Rat full
22	167	4.4	1596	7	ADE62326 Rat Prote
23	165	4.3	1596	2	AAW31347 Rat tumou
24	163	4.3	1500	8	ADM42998 Human INS
25	160.5	4.2	1562	4	ABG27493 Novel hum

26	160.5	4.2	1562	6	AAB32723	Aae32723 KIAA0322
27	160.5	4.2	1585	6	ABP97171	Abp97171 Human NED
28	160.5	4.2	1585	6	ABP98330	Abp98330 Amino aci
29	160.5	4.2	1607	4	AAM39192	Aam39192 Human pol
30	160.5	4.2	1664	4	AAM40978	Aam40978 Human pol
31	160.5	4.2	1702	4	ABG08505	Abg08505 Novel hum
32	158.5	4.2	5303	4	ABBE67866	Abbe67866 Drosophil
33	158	4.2	2768	4	ABBE68397	Abbe68397 Drosophil
34	157	4.1	897	8	AD116260	Ad116260 Human nuc
35	157	4.1	2766	8	ADF08473	Adf08473 Rat PAPIN
36	156.5	4.1	1454	7	ADD93813	Add93813 Human apo
37	156.5	4.1	1503	2	AAW48845	Aaw48845 Human rec
38	156.5	4.1	1503	7	ADP93815	Adp93815 Human HJ0
39	156.5	4.1	1503	8	ADP55733	Adp55733 Human PRO
40	156	4.1	888	4	ABG23910	Abg23910 Novel hum
41	156	4.1	888	4	ABG13903	Abg13903 Novel hum
42	154	4.0	2992	8	ADP30190	Adp30190 Human sec
43	154	4.0	3065	8	ADP30259	Adp30259 Human sec
44	154	4.0	4873	6	ABO14747	AbO14747 Novel hum
45	151	4.0	1090	6	ABP98840	Abp98840 Human str
46	151	4.0	1098	6	ABO07120	AbO07120 Novel hum
47	150.5	4.0	746	5	AD128089	Ad128089 Human nuc
48	150.5	4.0	764	5	AAO17706	Aao17706 Human PPA
49	150.5	4.0	1391	4	ABG08504	Abg08504 Novel hum
50	150	3.9	787	7	ADM05088	Adm05088 Human PRO
51	150	3.9	1062	3	AB40294	Ab40294 Human ORF
52	149.5	3.9	730	7	ADG10872	Adg10872 Human STA
53	149.5	3.9	969	4	AAU30116	Aau30116 Novel hum
54	149	3.9	1577	7	ADD71149	Add71149 Human int
55	149	3.9	1857	7	AAV53970	Aay53970 Human per
56	149	3.9	1857	3	ADJ70367	Adj70367 Human hea
57	149	3.9	1857	8	ADO19538	Ado19538 Human PRO
58	148.5	3.9	428	5	ABG93245	Abg93245 C. albica
59	148.5	3.9	553	7	ADM05820	Adm05820 Human pro
60	148.5	3.9	878	5	ABG91810	Abg91810 Human int
61	148.5	3.9	1192	4	ABBS9642	Abbs9642 Drosophil
62	148	3.9	1142	2	AAW81546	Aaw81546 Tumour re
63	148	3.9	1142	6	ABU08929	Abu08929 Human tum
64	148	3.9	1142	7	ADD25518	Add25518 Binding d
65	148	3.9	1142	7	ADG18025	Adg18025 MAGE-cl p
66	148	3.9	1142	7	ADM33107	Adm33107 Human tum
67	148	3.9	1142	8	ADI79399	Adi79399 Human MAG
68	148	3.9	1208	4	ABG24167	Abg24167 Novel hum
69	147.5	3.9	951	3	AAV58634	Aay58634 Protein r
70	147	3.9	1132	8	ADM42940	Adm42940 Human INS
71	147	3.9	3394	7	ADJ68723	Adj68723 Human hea
72	146	3.8	725	4	AAB95238	Ab95238 Human pro
73	145.5	3.8	1052	8	ADK71860	Adk71860 Human kin
74	145.5	3.8	2353	8	ADQ97927	Adq97927 Mouse can
75	145	3.8	614	4	AAB95229	Ab95229 Human pro
76	145	3.8	614	4	AAM94010	Aam94010 Human sto
77	145	3.8	614	8	ADR09871	Adr09871 Human pro
78	144.5	3.8	959	8	ADR10462	Adr10462 Human pro
79	144.5	3.8	1582	5	ADO85780	Ado85780 Rac/axin/
80	144	3.8	835	5	ABBS5681	Abbs5681 Human dif
81	144	3.8	835	8	ADRI4035	Adri4035 Human NF-
82	144	3.8	835	4	ABM80164	Abm80164 Tumour-as
83	143.5	3.8	1413	4	ABBG2045	Abbg2045 Drosophil
84	142.5	3.7	895	8	ABM83955	Abm83955 Human dia
85	142.5	3.7	1572	5	ABBS97562	Abbs97562 Novel hum
86	142.5	3.7	1572	6	ABP98339	Abp98339 Amino aci
87	142	3.7	406	5	ABG93152	Abg93152 S. cerevi
88	142	3.7	603	4	AAU30115	Aau30115 Novel hum
89	142	3.7	1455	4	AAW79120	Aam79120 Human pro
90	141.5	3.7	448	7	ADM06070	Adm06070 Human pro
91	141.5	3.7	1056	7	ADD18754	Add18754 Human dia
92	141.5	3.7	2194	4	AAM40114	Aam40114 Human POL
93	141.5	3.7	2724	6	ABP96961	Abp96961 Human BMC
94	141	3.7	993	2	AAV49897	Aay49897 Rat TAO2
95	141	3.7	993	8	ADJ58785	Adj58785 Rat TAO2
96	141	3.7	1911	4	ABM61038	Abm61038 Drosophil
97	140.5	3.7	1714	4	ABG02138	Abg02138 Novel hum
98	140	3.7	237	2	AAV66187	Aay66187 Human bla

99 140 3.7 685 8 ADR09563 Human pro
100 139.5 3.7 635 4 ABG22101 Novel hum

ALIGNMENTS

RESULT 1
ADQ88458
ID ADQ88458 standard; protein; 713 AA.
XX
AC ADQ88458;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human GADD34-like (GADD34L) protein.
XX
KW Oxidative stress; neuronal ischaemia; heart ischaemia; renal damage;
KW autoimmune disease; neurodegenerative disorder; therapy; GADD34-like;
KW GADD34L; eIF2alpha-specific regulatory subunit; phosphatase; human.
XX
OS Homo sapiens.
XX
PN US2004142345-A1.
XX
PD 22-JUL-2004.
XX
PF 28-AUG-2003; 2003US-00650482.
XX
PR 06-SEP-2002; 2002US-0408679P.
XX
PA (ROND/) RON D.
PA (JOUS/) JOUSSE C.
XX
PI Ron D, Jousse C;
XX
DR WPI; 2004-552556/53.
DR N-PSDB; ADQ88457.
XX
PT Screening test substances for preventing or treating disease involving
PT oxidative stress, by testing test substances for its ability to inhibit
PT activity of GADD34L and identifying test substance that inhibits activity
PT of GADD34L.
XX
PS Disclosure; SEQ ID NO 2; 30pp; English.
XX
CC The present invention relates to a method of screening several test
CC substances for preventing or treating diseases involving oxidative stress
CC such as neuronal ischaemia, heart ischaemia, renal damage induced by
CC ischaemia or toxins, autoimmune diseases and neurodegenerative disorders.
CC The method involves testing the test substances for its ability to
CC inhibit the activity of GADD34-like (GADD34L), also referred to as
CC eIF2alpha-specific regulatory subunit of phosphatase, and identifying the
CC test substance which inhibits the activity of GADD34L. The present
CC sequence is the human GADD34L protein.
XX
SQ Sequence 713 AA;

Query Match 100.0%; Score 3804; DB 8; Length 713;
Best Local Similarity 100.0%; Pred. No. 7.5e-303;
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPEGTGSRKRLGPRAGFRFPFPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60
DB 1 MPEGTGSRKRLGPRAGFRFPFPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60
QY 61 SYWTKLLSQLAPLPGLLQKVLINSQLPGMFPTRWLDVAGYISALRKALKREKPAAPTA 120
DB 61 SYWTKLLSQLAPLPGLLQKVLINSQLPGMFPTRWLDVAGYISALRKALKREKPAAPTA 120
QY 121 QKSLSSLDSSDPSVTSPLDWLEEGIHQVSPDLKLEKAGSALDPAQAFLLEQOL 180
DB 121 QKSLSSLDSSDPSVTSPLDWLEEGIHQVSPDLKLEKAGSALDPAQAFLLEQOL 180

QY 181 WGVLLPSSLSQRLYSNRELSSPSGLNLTQRIIDNFVSVSYLLNPSYLDPCPRLEVSQN 240
DB 181 WGVLLPSSLSQRLYSNRELSSPSGLNLTQRIIDNFVSVSYLLNPSYLDPCPRLEVSQN 240
QY 241 SDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWQCQCPPLSTEGTLPETIHLRMKR 300
DB 241 SDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWQCQCPPLSTEGTLPETIHLRMKR 300
QY 301 LEFLQOANKGQDLPTPDQNGYHSLREHSLRMDPKHCRDNPTQFVPAAGDIIFGNTOES 360
DB 301 LEFLQOANKGQDLPTPDQNGYHSLREHSLRMDPKHCRDNPTQFVPAAGDIIFGNTOES 360
QY 361 TEEKIELLTTEVPLALEEESPSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLFISARPA 420
DB 361 TEEKIELLTTEVPLALEEESPSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLFISARPA 420
QY 421 CSNKLDIYILGGASSDLETSDPEGEDWDEAEEDGDFDSSLSLSDSDLEQDPEGLHLWNS 480
DB 421 CSNKLDIYILGGASSDLETSDPEGEDWDEAEEDGDFDSSLSLSDSDLEQDPEGLHLWNS 480
QY 481 FCSVDYPNPNFTATIQTAARIIVPEEPSDSEKOLSGKSDLENSQSOSGLPETPHSSGEE 540
DB 481 FCSVDYPNPNFTATIQTAARIIVPEEPSDSEKOLSGKSDLENSQSOSGLPETPHSSGEE 540
QY 541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCROSKTPSESIVASE 600
DB 541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCROSKTPSESIVASE 600
QY 601 CHTLLSCKVQLLGSESECPDSVQORDVLSGGRHTRVKKVTFLEEVTEYVYISGDEDRKG 660
DB 601 CHTLLSCKVQLLGSESECPDSVQORDVLSGGRHTRVKKVTFLEEVTEYVYISGDEDRKG 660
QY 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNLQTCFKGLNVLKQC 713
DB 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNLQTCFKGLNVLKQC 713

RESULT 2
AAB95876
ID AAB95876 standard; protein; 713 AA.
XX
AC AAB95876;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18965.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length
 XX cDNAs.
 PS Claim 8; SEQ ID NO 18965; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AA92446 to AA95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 713 AA;

Query Match 99.7%; Score 3794; DB 4; Length 713;
 Best Local Similarity 99.7%; Pred. No. 5e-302;
 Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEGTGGSRKLGPRAGFRWPPPPRRSOGSSKFPETLGPENSGNPTLLSSAQPTRV 60
 DB 1 MEGTGGSRKLGPRAGFRWPPPPRRSOGSSKFPETLGPENSGNPTLLSSAQPTRV 60
 QY 61 SYWTKLSQLLAPLPGLLQKVLINSQFGGFFTRWLDFAQVYSGALRALKGREKPAAPTA 120
 DB 61 SYWTKLSQLLAPLPGLLQKVLINSQFGGFFTRWLDFAQVYSGALRALKGREKPAAPTA 120
 QY 121 QKSLSSQLQSSDSPSVTSLDWLBEHGHQVSPDPLKLELKAGSALDPAQAFLSQOL 180
 DB 121 QKSLSSQLQSSDSPSVTSLDWLBEHGHQVSPDPLKLELKAGSALDPAQAFLSQOL 180
 QY 181 WGVLELLSSLOSRLYSNRELSSPSGLNTORINFSVYLLNPSYLDLCPPLREVSQN 240
 DB 181 WGVLELLSSLOSRLYSNRELSSPSGLNTORINFSVYLLNPSYLDLCPPLREVSQN 240
 QY 241 SDGNSEVVGFOTLTPSSCLREDHCHPOPLSAELIPASWQCPLSTEGPLIEHILRMKR 300
 DB 241 SDGNSEVVGFOTLTPSSCLREDHCHPOPLSAELIPASWQCPLSTEGPLIEHILRMKR 300
 QY 301 LEFLQANKGQDLPTPDQDNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
 DB 301 LEFLQANKGQDLPTPDQDNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
 QY 361 TEKIELLTTRVPLALEEESSECPSSSEIPEWKEPGEGRISVVYDYSYLEGDLPI SARPA 420
 DB 361 TEKIELLTTRVPLALEEESSECPSSSEIPEWKEPGEGRISVVYDYSYLEGDLPI SARPA 420
 QY 421 CSNKLIDYILGASDLTSSDPGEDWDEAEADGDFDSSLSLSDLEQDPPEGLHLWNS 480
 DB 421 CSNKLIDYILGASDLTSSDPGEDWDEAEADGDFDSSLSLSDLEQDPPEGLHLWNS 480
 QY 481 FCSVDPPNPQNTATIQTAARIVPEEPSDSKGLSGKSDLENSSQSGSLPTPEHSGEE 540
 DB 481 FCSVDPPNPQNTATIQTAARIVPEEPSDSKGLSGKSDLENSSQSGSLPTPEHSGEE 540
 QY 541 DDWESSADEAESLKLWNSFCNSDDPPYNPLNFKAPQTSGENEKCRCRSTPESIVASE 600
 DB 541 DDWESSADEAESLKLWNSFCNSDDPPYNPLNFKAPQTSGENEKCRCRSTPESIVASE 600

QY 601 CHTLLSCKVQLLGSQSECPDSVQRDVLSGGRHTRHVKRKKVTFLEEVTEYIISGDEDRKG 660
 DB 601 CHTLLSCKVQLLGSQSECPDSVQRDVLSGGRHTRHVKRKKVTFLEEVTEYIISGDEDRKG 660
 QY 661 PWEEFARDGCRFQKRIQETEDATGYCLTTEHRRMFNRLOQCTCFKGLNLVKQC 713
 DB 661 PWEEFARDGCRFQKRIQETEDATGYCLTTEHRRMFNRLOQCTCFKGLNLVKQC 713
 RESULT 3
 AAB53401
 ID AAB53401 standard; protein; 707 AA.
 XX
 AC AAB53401;
 XX
 DT 09-MAR-2001 (first entry)
 DE Human colon cancer antigen protein sequence SEQ ID NO:941.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005883.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587534/55.
 DR N-PSDB; AAC98158.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer.
 XX
 PS Claim 11; Page 1503-1506; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins may
 CC also be used to prevent diseases such as neural disorders, immune system
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal
 CC disorders, wounds, renal disorders, infectious diseases, and
 CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
 CC sequences used in the exemplification of the present invention
 XX
 SQ Sequence 707 AA;

Query Match 95.3%; Score 3627; DB 3; Length 707;
 Best Local Similarity 98.8%; Pred. No. 2.5e-288;
 Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 584 GHHTLLSCKAQLLESQDNPCGGLGALAGERVTHIKRKKVTFLEEVTEYISGDEDRK 643
 Qy 660 GPWEEFARDGCRFKRIQETEDAIGYCLTFEHRERMFNRLO 700
 Db 644 GPWEEFARDGCRFKRIQETEAIGYCLAFEHRERKMFNRLR 684

RESULT 5
 AAM25448
 ID AAM25448 standard; protein; 153 AA.
 AC AAM25448;
 XX 16-OCT-2001 (first entry)
 XX Human protein sequence SEQ ID NO:963.

XX Human; cancer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiagregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression; disorder;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.
 OS
 XX W0200153455-A2.
 XX 26-JUL-2001.
 XX 22-DEC-2000; 2000WO-US035017.
 XX 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-457603/49.
 DR N-PSDB; AAM99389.
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
 XX Claim 20; Page 204; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and cells
 CC they are expressed in, such as: antinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnary;
 CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders
 XX Sequence 153 AA;
 SQ

Query Match 21.4%; Score 813; DB 4; Length 153;
 Best Local Similarity 100.0%; Pred. No. 2.8e-58;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 DPYNPLNFAPFQTSGENEKGCRDSKTPSESIVAISECHTLISCKVQLLSQSECPDSV 623
 Db 4 DPYNPLNFAPFQTSGENEKGCRDSKTPSESIVAISECHTLISCKVQLLSQSECPDSV 63
 Qy 624 QRDVLSGGRHTHVKKKVTFLIEVTEYISGDEDRKGPWEEFARDGCRFKRIQETEDAI 683
 Db 64 QRDVLSGGRHTHVKKKVTFLIEVTEYISGDEDRKGPWEEFARDGCRFKRIQETEDAI 123
 Qy 684 GYCLTFEHRERMFNRLOQTCFKGLNVLKQC 713
 Db 124 GYCLTFEHRERMFNRLOQTCFKGLNVLKQC 153

RESULT 6
 AAY84366
 ID AAY84366 standard; protein; 674 AA.
 XX AAY84366;
 AC AAY84366;
 XX 12-JUL-2000 (first entry)
 XX Amino acid sequence of a human apoptosis associated protein (GADD34).
 DE Human; apoptosis associated protein; GADD34; antiviral agent;
 KW growth arrest and DNA damage-inducible gene 34; viral infection;
 KW serine/threonine phosphatase.
 XX Homo sapiens.
 OS
 XX GB2342716-A.
 XX 19-APR-2000.
 XX 27-JUL-1999; 99GB-00017631.
 XX 14-SEP-1998; 98GB-00020025.
 XX (PFIZ) PFIZER LTD.
 XX Ciaramella G;
 XX WPI; 2000-285397/25.
 DR N-PSDB; AAZ99766.
 XX Screening assay for potential antiviral agents that modify
 PT serine/threonine phosphatase activity for treating human viral
 PT infections.
 XX Claim 6; Page 99; 106pp; English.

XX The present sequence represents a human apoptosis associated protein
 CC (GADD34). It is encoded by a growth arrest and DNA damage-inducible gene
 CC 34. It is used in an assay method for identifying an anti-viral agent
 CC that can affect the activity or expression of a nucleotide sequence or
 CC its expression product. The assay method comprises contacting an agent
 CC with a nucleotide sequence coding for a serine/threonine phosphatase or
 CC its expression product, and determining whether the agent affects the
 CC activity or expression of the nucleotide sequence. The assay is used for
 CC screening for antiviral agents useful in the treatment of human viral
 CC infections. The anti-viral modulators of type 1 protein phosphatase
 CC activity inhibit the action of a host component which interferes with the


```

227 YLDCPRLREVSQNSDGNSEVVGFOTLTPBSSCLREDH-CHPQPLSAELIPASWQCQPPPL 285
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 HRECCPAVE---EEDD-----EAVKKEAHTSTALSQSGSPSTWVSCP-- 224
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
286 STEGLPEIHHLRMKLEFLQOANKQDLP,TPDQNGYHSLLEEHSLLRMPKPCRONPTQ 345
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 ---GEENQATEKKTERTSKGARK-----TSVPSRSGSDPRS 259
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
346 FVPAAGDIPGNTQBSTEEKIELLTTEVPLALEESPBGCGPSSE-- -TPMEKEPG 397
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
260 WEYRSG-----EASEEKEEKAHKETGKGAAAPQPOSSAPAPRPOLKSWWCQPSDEEG 312
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
398 EGRISVYDYSILEGDLPI,ISARPACSNKLDIYILGASSDLETSSDP----- 443
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
313 EVK-----ALGAAEKGGEAECP,CP,PPPSAFLKAWVY 344
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
444 -EGEDWDEAEDDGGPDSOSSLSDSLOQDPG-----LHLNWSFCVDPYNPQNP 492
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
345 WPGEDTEEB-EDBEEDSDSGSDEEGEAEASSTPATGVFLKSW-----VYOPGED 396
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
493 TATIO,TAARIVPEEPSDEKOLSGKSDLENSQSGSLP-----ETPEH-- 535
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 T-----ESEEDEDSTGSAEDEREAE,TSASTPPASAFUKAWVYRPGEDTEEBED 445
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
536 ---SSGEDDWESSADRAESLUKNSFCNSDDDPYNPLNFKAPQTSG-----ENEGK 584
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
446 EDVDSDEKDDSEALGEAES-----DPHFSHP-DQRAHPRGMGYRPGKETEBEEA 495
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
585 CRD--SKTPSEGISV,ISECHTLLSCVKVLLGSQ-----ESEC,PD,SVQRDLVSGGRH,TH- 635
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
496 AEDWGEAEC,PF,RAV,-----YVPEKE,PP,PPWAP,PL,RL,OR,LR,KP,ET,PTH, 544
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
636 -----VKRK,KV,TF,LEE,VE,YYIS-----GDEDRKGPWEEF,ARD,CGCR,FQK,RI,OT,EDA,IG 684
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
545 POPETPLKARKV,FESEKV,TVH,FLAVWAGPAQAARQGEWQLARD,SR,SR,FARR,ITQ,QA,BEELS 604
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
685 YCLTFEHRERMF,NR,LOGTC,FKGL,NVL,KQ 712
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
605 PCLTTPAARARAWARL,NP,PLAP,IPALTO 632
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESIST. 8

RESULI 8	
ADD45119	
ID	ADD45119 standard; protein; 674 AA.
XX	
AC	ADD45119;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Human Protein XP_009097, SEQ ID NO 10552.
XX	
KW	Human; pain; neuronal tissue; gene therapy;
KW	spinal segmental nerve injury; chronic constriction injury; CCI;
KW	spared nerve injury; SNI; Chung.
XX	
OS	Homo sapiens.
XX	
PN	WO2003016475-A2.
XX	
PD	27-FEB-2003.
XX	
PF	14-AUG-2002; 2002WO-US025765.
XX	
PR	14-AUG-2001; 2001US-0312147P.
PR	01-NOV-2001; 2001US-0346382P.
PR	26-NOV-2001; 2001US-0333347P.
XX	
PA	(GEO) GEN HOSPITAL CORP.
PA	(FARB) BAYER AG.
XX	
PI	Woelf C, D'urso D, Befort K, Costigan M;
XX	

QY 493 TATIQTAAIRVPERPDSSEKDLGKSLNSQSGSLP-----ETPEH-- 535
 DB 397 T-----EEDDEDSTGSAEDREAEASSTPPASAFKAWYRPGEDTEBED 445
 QY 536 -----SSGEDDWESSADAEKSLKLNWSCNDDPYNPLNFKAPQTSG-----ENEKG 584
 DB 446 EDVSDKEDDSEALGEAES-----DPHPSHP-DQRAHFRGMYRPGKETEBEEA 495
 QY 585 CRD-SKTPSESVIAISECHTLSSCKVQLGSG-----ESECDSVQDVLSCGRH-- 635
 DB 496 AEDWGEAEPCFPRVAI-----YVGEKPPPPWAPPRLLRLQRLKRPTPTH 544
 QY 636 -----VRRKVTFLBEVTEYIS-----GDEDRKGPWEEFARDGCFQRIQETEDAIG 684
 DB 545 PDPTPLKARKVRSEKVTYVHFLAVWAGPAQAARQGPEQLARDRSFARRITQAEELS 604
 QY 685 YCLTFEHRERMFNRLOQTCFKGLNLVKQ 712
 DB 605 PCLTPAARARAWARLNPPPLAPALATQ 632

RESULT 9
 ADE54744
 ID ADE54744 standard; protein; 674 AA.
 XX
 AC ADE54744;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human Protein XP_009097, SEQ ID NO 549.
 XX
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX
 OS Homo sapiens.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 PA (GHEO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 DR WPI; 2003-268312/26.
 DR GENBANK; XP_009097.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 FS Claim 1; Page; 1017pp; English.
 XX

The invention discloses a composition comprising two or more isolated rat
 or human polynucleotides or a polynucleotide which represents a fragment,
 derivative or allelic variation of the nucleic acid sequence. Also
 claimed are a vector comprising the novel polynucleotide, a host cell
 which is differentially regulated in an animal subjected to pain and a
 kit to perform the method, an array, a method for identifying an agent
 that increases or decreases the expression of the polynucleotide sequence
 that is differentially expressed in neuronal tissue of a first animal
 subjected to pain, a method for identifying a compound which regulates
 the expression of a polynucleotide sequence which is differentially
 expressed in an animal subjected to pain, a method for identifying a

compound that regulates the activity of one or more of the
 polynucleotides, a method for producing a pharmaceutical composition, a
 method for identifying a compound or small molecule that regulates the
 activity in an animal of one or more of the polypeptides given in the
 specification, a method for identifying a compound useful in treating
 pain and a pharmaceutical composition comprising the one or more
 polypeptides or their antibodies. The polynucleotide or the compound that
 modulates its activity is useful for preparing a medicament for treating
 pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 therapy). The sequence presented is a human protein (shown in Table 2 of
 the specification) which is differentially expressed during pain. Note:
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic form directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 674 AA;
 SQ

Query Match 6.8%; Score 260.5; DB 7; Length 674;
 Best Local Similarity 21.9%; Pred. No. 6.3e-12;
 Matches 164; Conservative 80; Mismatches 263; Indels 241; Gaps 32;

QY 70 LLAPLGLLQKVLWSQFCGMPEPTR-WLDFAGVYSNL--RALKGR----- 112
 DB 21 LLSFVWGLLSRA--WSRL-RGLGLPELWVEAVKGAALVEAGLEGEARTPLAIPHTWGR 77
 QY 113 -----EKPAATAQKSLSSQLDSSDPVSTPLDMLDEEIGHWQYSPDLKLELKAKGSA 166
 DB 78 RPEEAEDSCGPGEDRETGLKTSLSLPEAWGLLD-DDDGMGEREATSVP---RGQSG 133
 QY 167 LDPAQAFLLEQQLWGVLLPSSLQRLYNNRELGGSPSGPLNRIQDNFVSVYLNPS 226
 DB 134 FADQORAPL-----SPSLLIRTLOGSDKNPGEKAEEGVAEEGVNKS-----YPPS 182
 QY 227 YLCFPRLEVSYQNSDGNSEVGFQTLTPRESSCLREDH-CHPQPLSAELIPASWQGCPL 285
 DB 183 HRECCPAVE---BEDD-----BEAVKGAHRTSTALSFGSKPESTWVSCP-- 224
 QY 286 STEGLPEIHLMKRLFLQANKQDLPTPDQDNGYHSLDEEHSLLRMDPKHCRDNPQTQ 345
 DB 225 ---GEENQATEDKRTSKGARK-----TSVPSRSGSDPS 259
 QY 346 FVPAAGDIPGNTOESTEEKIELTTEVPLALEESPSGCPSSP-----IPMEKEPG 397
 DB 260 WEYRSG-----EASEKEKAHKETGKEAAGPPOSSAPAPQPLKSWWCQPSDEEG 312
 QY 398 EGRISVVDYSYLSGDLPI SARPACSNKLDIYILGGASSDLETSSDP----- 443
 DB 313 EVK-----ALGAAEKDGEAECPCIPPPSAFLKAWY 344
 QY 444 -EGEDMDEAEDDGFDSLSLSDLDLQDDEG-----LHLMNSFCSDVPYNQNF 492
 DB 345 WPGEDTEEE-EDDEEDSDSGDEEGEAASSTPATGVFLKSW-----VYQGED 396
 QY 493 TATIQTAAIRVPERPDSSEKDLGKSLNSQSGSLP-----ETPEH-- 535
 DB 397 T-----EEDDEDSTGSAEDREAEASSTPPASAFKAWYRPGEDTEBED 445
 QY 536 -----SSGEDDWESSADAEKSLKLNWSCNDDPYNPLNFKAPQTSG-----ENEKG 584
 DB 446 EDVSDKEDDSEALGEAES-----DPHPSHP-DQRAHFRGMYRPGKETEBEEA 495
 QY 585 CRD-SKTPSESVIAISECHTLSSCKVQLGSG-----ESECDSVQDVLSCGRH-- 635
 DB 496 AEDWGEAEPCFPRVAI-----YVGEKPPPPWAPPRLLRLQRLKRPTPTH 544
 QY 636 -----VRRKVTFLBEVTEYIS-----GDEDRKGPWEEFARDGCFQRIQETEDAIG 684
 DB 545 PDPTPLKARKVRSEKVTYVHFLAVWAGPAQAARQGPEQLARDRSFARRITQAEELS 604
 QY 685 YCLTFEHRERMFNRLOQTCFKGLNLVKQ 712
 DB 605 PCLTPAARARAWARLNPPPLAPALATQ 632

autoimmune disorder; hyper immune activity;
 aberrant acute phase response; hypercongenital condition; birth defect;
 necrotic lesion; wound; organ transplant rejection;
 aberrant signal transduction; proliferating disorder; cancer;
 HIV propagation; human.
 Homo sapiens.
 WO2004065577-A2.
 05-AUG-2004.
 13-JAN-2004; 2004WO-US000798.
 14-JAN-2003; 2003US-0440068P.
 12-MAY-2003; 2003US-046957P.
 (BRIM) BRISTOL-MYERS SQUIBB CO.
 Nadler SG, Neubauer MG, Feder JN, Carman J;
 WPI; 2004-562168/54.
 N-PSDB; ADR14088.
 New isolated polynucleotides and polypeptides associated with NF-kappaB pathway, useful for diagnosing, treating, or preventing disorders or diseases associated with NF-kappaB pathway.
 Claim 6; SEQ ID NO 90; 237pp; English.
 This invention relates to the novel association of protein sequences (and the genes which encode them) to the NF-kappaB pathway. The invention may be useful for the production of compounds with an anti-inflammatory, cytosolic, hepatotropic, virucide, antiarthritic, antirheumatic, gastrointestinal-gen, antasthmatic, antiarteriosclerotic, immunomodulator, cerebroprotective, vasotropic, immunosuppressive or vulnary activity or for gene therapy. The proteins and nucleotides are useful for diagnosing, preventing, treating, or ameliorating conditions or diseases associated with the NF-kappaB pathway. The condition is an immune disorder, an inflammatory disorder, an inflammatory disorder related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell survival, evasion of immune responses, rheumatoid arthritis, inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper immune activity, disorders related to aberrant acute phase responses, hypercongenital conditions, birth defects, necrotic lesions, wounds, organ transplant rejection, conditions related to organ transplant rejection, disorders related to aberrant signal transduction, proliferating disorders, cancers and HIV propagation in cells infected with other viruses. The present sequence is that of a human protein which is subject to the novel association with the NF-kappaB pathway of the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from Genbank.

Query Match 6.8%; Score 259.5; DB 8; Length 674;
 Best Local Similarity 22.3%; Pred. No. 7.7e-12;
 Matches 166; Conservative 82; Mismatches 261; Indels 237; Gaps 34;
 70 LLAPLGLLQKLVLSQIFGGMFTR-WLDFAGVYSAL--RALKGR----- 112
 21 LLSPVGMGLLSRA--WSRL-RLGLPLPLVAVGALVEAGLEGEARTPLAIPHTPWGR 77
 113 -----EKPAAPTAQKLSLSQLSDPSVTSPLDLWLEIGHWQYSPDLKLELKAGSA 166
 78 RPEGEADSGGFGEDRTILGLTKTSSSLPEAWGLLD-DDDGMYGREATSVP--RGQGSQ 133
 167 LDPAQAFLLEQLMGVELLPSLSQSLRYNRELSSPSGPLNTQRIIDNFSWVSYLLNPS 226

Db 134 FADQORAPL-----SPSLLIRTLQSGDKNPGKEKAEEGVVAEEGVNKFS-----YPPS 182
 Qy 227 YLDCFPRLVSYQNSDGNSEVGFQTLTPRESSCLREDH-CHPQLSLAELIPASWQGCPL 285
 Db 183 HRECCPAVE---BEDD-----BEAVKKEAHTSTALSFGSKPSTWVSCP-- 224
 Qy 286 STEGLPEIHLMKRLFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPQ 345
 Db 225 ---GEENQATEDKRTYRSKARK-----TSVPRSSGSDPRS 259
 Qy 346 FVPAAGDIPGNTQESTEEKIELLTTEVPLALEBESPSEGC--PSSEIPMEK-----E 395
 Db 260 WEYRSG-----EASEKEE-----KAHEETGKEAAPGPOSSAPAPQRPOLKSWCQ 305
 Qy 396 PGEGRISVVDYSYLEGLDPLISARPAKSNKLIDYLGASDSDLETSSDP----- 443
 Db 306 PSDEESEV-----KP-----LGAAEKDGAEACPPCIPPPSAFLKAW 342
 Qy 444 ---EGEDWDEEAEDDGFDSLSLSDLEODPEG-----LHLWNSFCSVDYPNPQ 490
 Db 343 VYWPGEDEEE-EDDEDEDSGSDSEEGEAASSSTPATGVFLKSW-----VYQPG 394
 Qy 491 NPTATIQTAAARIPEBPSDESKDLGKSDLENSQSGSLP-----ETPEH 535
 Db 395 EDT-----EEEEDESDTGSADEREAETSASTPPASAFKAWVVRPGEDTEEE 443
 Qy 536 -----SSGREDWESSADEAESIKLWNSFCNSDDPYNPLNFKAPFOTSGENEGKCRD 589
 Db 444 EDEDVSDKEDDSEALGAEES-----DHPH-----SHPDQSAHFRGWGYRPGK 488
 Qy 590 TPSESIVA-----ISECHTLLSCKVQLLGSQ-----ESECPDSVQRDLVLSGGRH 635
 Db 489 ETEEEEAEDWGEAEPFPRVA--IYVGEKPPPPWAPPPLPLRLQRLKRPETPTDPD 546
 Qy 636 ----VRKKVTFLEVTVEYVIS-----GDEDRKPWFEEFARDGCRFOKRIQETEDALGYC 686
 Db 547 PETPLKARKVRFSEKVTVHFLAVWAGPAQAARQGPWEQLARDRSFRFARRIAQAQEELSPC 605
 Qy 687 LTFEHRERMFENRLOGICFKGLNLVKQ 712
 Db 607 LTPAARARAWARLUNPPLAPALPTQ 632
 RESULT 12
 AAW99891
 ID AAW99891 standard; protein; 674 AA.
 AC AAW99891;
 DT 09-JUN-1999 (first entry)
 DE Human growth regulator protein GRREG.
 KW Human; growth regulator protein; GRREG; cancer.
 OS Homo sapiens.
 PN W09902680-A1.
 PD 21-JAN-1999.
 PF 30-JUN-1998; 98WO-US013409.
 PR 11-JUL-1997; 97US-00893852.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Lal P, Shah P, Corley NC;
 DR WPI; 1999-120881/10.
 DR N-PSDB; AAX19798.
 XX

PT Novel human growth regulator protein - useful in the treatment of
PT cancers.
XX
XX Claim 1; Fig 1; 70pp; English.
PS
PS
XX
CC The present sequence represents a human growth regulator protein,
CC designated GRREG. The protein can be used in methods to treat cancer. The
CC protein can also be used in binding assays to detect (ant)agonists of
CC GRREG. A fragment from the nucleic acid sequence encoding GRREG can be
CC used as a probe for detecting GRREG encoding sequences (especially in PCR
CC amplified samples)
XX
XX Sequence 674 AA;
SQ

Query Match	6.6%	Score	252.5;	DB 2;	Length	674;
Best Local Similarity	21.9%;	Pred. No.	2.9e-11;			
Matches	164;	Conservative	80;	Mismatches	262;	
				Indels	243;	Gaps
					32;	

Qy	70	LLAPLGLLOKVLWSQLFGOMFTR--WLDFAGVYSAL---RALKGR-----	112
Db	21	LLSPVMSLLSRA--WSRL-RGLGPLEPLWLVAVKGAALVEAGLEGEARTPLAIPHTPWGR	77
Qy	113	-----EKPAAPTAKSLSSILQLODSSDPVSTPDLMDLEEGIHWQYSPDLKLELKAKGSA	166
Db	78	RPEEAEADSGGFGEDRETGLUKTSSSLPEANGLLD-DDDGMYGKEREATVSP---RGQSGS	133
Qy	167	LDPAQAQFLLEQQLWGVLELLPSSILQSRILSYNRELGSSPGPLNIQRIDNFVSVYLLNPS	226
Db	134	PADQORAPL-----SPSLLTRTLQSGDKNFGEEKABEGVAEEGVNKFS-----YPS	182
Qy	227	YLDGCPRLVSYQNSDGNSEVVGFTLLTPESCLREDH-CHPQPLSABELIPASWOGC9PPL	285
Db	183	HRECCPAVE---EBDD-----EBAVKKEAHRSTLSALSPGSKPSTWVSCP---	224
Qy	286	STGCLPEIHILRMKRLSFLOAKNGQDLPTPDQDNGVHSLSEHSLLRMDPKHCDNPTQ	345
Db	225	---GREENQATEDKRTERSKGARK-----TSVSRSSGSDPRS	259
Qy	346	FVPAAGDIPIGNTQBSTBEKIELLATTEVPLALEEBSPEGCPSPSEIPMEKEPGEGRISVVD	405
Db	260	WEYRSG-----EASEEKEE-----KAHEETCKGEAAPG-----	287
Qy	406	YSYLEGDLPTISARPACSNKLI.DY-----ILGAGSDLETSSDP-----	443
Db	288	-----POSSAPAQRPQLKSNWCQCPDEEESYKALGAAEKDGEACPCIPPPSAFL	339
Qy	444	-----EGEDWDEAEDDGFSDSLSDLEQQDEG-----LHLNWSFCSVDPY	487
Db	340	KAWYWPGEUTEER-EDEEDEDSDSGDEEGEAEASSTPATGVFLKSW-----VY	391
Qy	488	NPQNFATTIQAARIVPEP9DSEKDLGSKGDLNSENSQSGSLP-----ET	532
Db	392	QPGEDT-----EEDEEDSDTGAEDEREAEATSSTPPASAFKAWVYRPGEDT	440
Qy	533	PEH-----SSGEEDDWESSADAESLKLWN5FCNSDDPYNPLNPKAFQPTSGENEGKGR	586
Db	441	EEEBDEDVDSDEKEDDSEALGAEAS-----DHP-----SHPDQSAHFAGWGRV	485
Qy	587	DKTTPSESIVA-----ISECHTLLSCKVQLLGSQ-----ESECPDSVQORVLSGGRH	635
Db	486	PGKETEBEEAAEDNCEAEPC9FRVA--IYVPGKEPPPPWAPPLRLPLRLQRLKRRPETPTH	543
Qy	636	-----VKRKVVFLFEVTEBYI5-----GDEDRKGPEWBFARDCGRFKRIQETDEAI	683
Db	544	DPDPETPLKARKVRFSEKVYVHFLAVWAGPAQAARQGWELARD9SFARRITQAEEL	603
Qy	684	GCYLTFEHRERMFNRLOQTCFKGLNVLKQ	712
Db	604	SPCILTPAARAWARLRRNPPLAPITPALTO	632

RESULT 13
AAAY41102

ID	AAV41102 standard; protein; 657 AA.
XX	
AC	AAY41102;
XX	
DT	17-JAN-2000 (first entry)
XX	
DE	Rat MyD116 protein sequence.
XX	
KW	Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW	cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;
KW	MyD116 protein.
XX	
OS	Rattus sp.
XX	
PN	WO949898-A1.
XX	
PD	07-OCT-1999.
XX	
FF	31-MAR-1999; 99WO-US007199.
XX	
PR	31-MAR-1998; 98US-00052753.
XX	
PA	(UYCO) UNIV COLUMBIA NEW YORK.
XX	
PI	Fisher PB;
XX	
DR	WPI; 1999-591184/50.
XX	
PT	Novel vectors useful for studying the progression of cancer.
XX	
PS	Disclosure; Fig 3A-B; 25ipp; English.
CC	The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC	gene) regulatory region functionally linked to a gene encoding a product
CC	that causes or may be induced to cause the death or inhibition of cancer
CC	cell growth. A vector of the invention which contains a gene encoding
CC	thymidine kinase or a product which causes the cell to express a specific
CC	antigen can be administered along with gancyclovir or acyclovir, or an
CC	antibody or fragment to the antigen, respectively, to treat cancer in a
CC	subject. The PEG-3 gene is useful for generating new cloning and
CC	expression vectors, transfected cells, and for developing methods for
CC	cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC	a source of primers and probes to study the progression of cancer, and to
CC	detect the presence of the gene. The present sequence represents the
XX	amino acid sequence of the rat MyD116 protein
XX	
SQ	Sequence 657 AA;
Query Match 6.3%; Score 238; DB 2; Length 657;	
Best Local Similarity 22.4%; Pred. No. 4.3e-10;	
Matches 173; Conservative 81; Mismatches 233; Indels 284; Gaps 38;	
QY	53 SAQPETRVSYWTXLLS-QLLAPLPGILQKVLWSQLFGGMFTRFLDPAGVYSALRALKG 111 : : : : : : : :
Dd	4 SPRPQ-HVLHRDAHNFYLLSPLMGLSRA--WSRLRGPEVPFAWL----- 46
QY	112 REXPAAPTAKSLSSQLDDSDPVTSPLDWLEEGIHQWYSPDLKLKELKAGSALDPAA 171 : : : : : : : :
Dd	47 -----AKTVGADQIE-----AKTVGADQIE-----A 58
QY	172 QAFLEQLQMGVELLPSSLQSRLYSNRELSSPSGPLNTQRIDNFVSVYLNPNSYLDGF 231 : : : : : : : :
Dd	59 AALLTTPVSG-NLLP-----HGTEESGSPEQSQAAQRL-----CL 94
QY	232 PRLEVYSQNSDGNSEVVGFTLTTPESSCLRCHDCHPOPLSAEILPASWGQCPLSTGLP 291 : : : : : : : :
Dd	95 VEAESSPPETWGLSNVDYNA-KPGODDLREKEMERTAGKATLQAPAGLOAD----- 145
QY	292 EIIHLRWKLEFLQQANG--QDLTPPDNDNGYHSLEEHSLR-----MDPKHCR 340 : : : : : : : :
Dd	146 -----KRIGEVVAAREGVAEPAYTTSLEGPAINEEDGETVKTYQASAAIAFGYKP 198
QY	341 DNPTQFVPAAGDIPGNTQSBTEBKIELLTITVEFLALEEESPGECPSS--EIPMEKEPGE 398

Db 199 STPVFLGEA-----EQATEEK-----GTE-NKADPSNPSGSHSRAWEYYSREKPKQ 247
 Qy 399 GRISVVDVSYLGLDLPISARPA-----CSNKLIDVILGASDLTSDP-----EG 445
 Db 248 -----EGEAKVEAHRAQOGHPCRNABAE--GGPETTFVCTGNFLKAWVYRPG 294
 Qy 446 EDWDEE-----AEDD-----GFSDSLSDS-LEQDPEGL- 475
 Db 295 ETEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTEEDSDSAEEDTAGT 354
 Qy 476 ---HLWNSFCSDVPYNPQNT-----ATQTAARI---VPEEPSDE 511
 Db 355 ATPHT-SAFKAWVYRPGEDTEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTE 413
 Qy 512 KDLSGKSDL-----ENSSQSGSLPET-----PEHSSGEEDWES-----S 546
 Db 414 EB---NSDLSAEEDTAQTGATHTSPFLKAWVYRPGEDTEEDTEEDSENAPGDS 470
 Qy 547 ADEAESKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDSTPSESIVAI 598
 Db 471 ADSSQSPCLQPCORCLPGEKTKRGEPP--PL-FQVAFYLPGEK-----PESWAA 517
 Qy 599 SECHTLCKVOLLGS-----QSECEPDSVQRDLVSGGRHTHVKRKKVTFLEEVTEYIS- 653
 Db 518 PKLPLRLQRLRLFKAPTRDQDPEIP-----LKARKVHFAEKTVHFLAV 562
 Qy 654 ----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERFNLQ 700
 Db 563 WAGPAQAARRGPEQFARDRSRFAARIAQAEKLGPLYLTPDSRARAWARL 613

RESULT 14

ABBS7373
 ID ABBS7373 standard; protein; 657 AA.

AC ABBS7373;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:1062.

KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

PN WO200188188-A2.

PD 22-NOV-2001.

PF 18-MAY-2001; 2001WO-JP004192.

PR 18-MAY-2000; 2000JP-00145977.

PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

DR N-PSDB; ABI99909.

PT Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.

PS Claim 2; Page 2676-2679; 2690pp; English.

CC The present invention describes a method for examining ischaemic

CC conditions, comprising measuring the expression levels of particular

CC genes (I) in a test sample or determining the expression profile of a

CC gene group in the sample comprising genes selected from (I). The method

CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI9912, encoding the
 CC protein sequences in ABBS7020 to ABBS7374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention

XX SQ Sequence 657 AA;

Query Match 6.2%; Score 237; DB 5; Length 657;

Best Local Similarity 21.8%; Pred. No. 5.2e-10;

Matches 169; Conservative 81; Mismatches 233; Indels 292; Gaps 35;

Qy 53 SAQPETRVSVYTKLLS-QLLAPLPLGLKQVILVNSQLFGMPTRWLDVAGVYSALRAK 111

Db 4 SPRPQ-HVLHWRDAHNFYLLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46

Qy 112 REXPAAPTAQKSLSSLDSDSPVTSPLDWLEBEGITHWQYSPDLKLEKAKGSALDPA 171

Db 47 -----AKVTGADOIE-----A 58

Qy 172 QAFLEQQLWGVLLPSSLSQSLRYSNRELSSPGPLNIQRIIDNFSVYLLNPSYLD 231

Db 59 AALLTPFVSG-NLLP-----HGETEESGSEQSOAAQRL-----CL 94

Qy 232 PRLEVSQNSDGNSEVVGFTLTPESCLREDHCHPOPLSAELIPASWQCPPLSTGLP 291

Db 95 VEAESSPPETWGLSNVDEYNA-KPGQDDLREKEMERTAGKATLQAPAGLQAD----- 145

Qy 292 EIHHLRMKRLFLQOANKG---QDLTPDQDNGVHSLSEESLLR-----MDPKHCR 340

Db 146 -----KRLGEVVAEEGVAEPAYPTSQLEGGPAENEEDGETVTKTQOASASTAPYKP 198

Qy 341 DNPTQFVPAAGDIPGNTQESTEEKIELLTTVEPLALEEESPSGCPSS--EIPWEKEPGE 398

Db 199 STPVFLGEA-----EQATEEK-----GTE-NKADPSNPSGSHSRAWEYYSREKPKQ 247

Qy 399 GRISVVDVSYLGLDLPISARPA-----CSNKLID-----YILGAS 434

Db 248 -----EGEAKVEAHRAQOGHPCRNABAEEGGPEFTTCTGNFLKAWVYRPGED 296

Qy 435 SDLETSSDPE-----GEDWDEAEDDGPDSLSLSDDL 468

Db 297 TEEDNSDSAEEDTAQTGATHTSAFLKAWVYRPGEDTEER-----DSDSDSEEDT 350

Qy 469 EQDPEGLHLWNSFCSDVPYNPQNT-----ATQTAARI---VPEEP 507

Db 351 AQTGATPHT-SAFKAWVYRPGEDTEEDNSDLSAEEDTAQTGATHTSAFLKAWVYRPG 409

Qy 508 SDSEKDLGKSDL-----ENSSQSGSLPET-----PEHSSGEEDWES----- 545

Db 410 EDTEEE---NSDLSAEEDTAQTGATHTSPFLKAWVYRPGEDTEEDTEEDSENAPG 466

Qy 546 ---SADEAESKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDSTPSES 594

Db 467 DSETADSSQSPCLQPCORCLPGEKTKRGEPP--PL-FQVAFYLPGEK-----PES 513

Qy 595 IVAISECHTLCKVOLLGS-----QSECEPDSVQRDLVSGGRHTHVKRKKVTFLEEVTEY 650

Db 514 PWAAPKLPLRLQRLRLFKAPTRDQDPEIP-----LKARKVHFAEKTVH 558

Qy 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERFNLQ 700

Db 559 FLAVWAGPAQAARRGPEQFARDRSRFAARIAQAEKLGPLYLTPDSRARAWARL 613

RESULT 15

AAW79958

ID AAW79958 standard; protein; 578 AA.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 09:17:12 ; Search time 19.202 Seconds
(without alignments)
2771.838 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEFPTGGSRKRLGPRAGFRF.....RMFNRLQGTCFKGLNLVKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	252.5	6.6	674	3	US-08-893-852A-1
2	237	6.2	657	3	US-08-893-852A-3
3	236	6.2	657	3	US-08-821-818-3
4	236	6.2	657	4	US-09-052-753B-3
5	219	5.8	578	4	US-09-052-753B-7
6	201	5.3	590	3	US-08-893-852A-4
7	201	5.3	590	3	US-08-821-818-2
8	201	5.3	590	4	US-09-052-753B-2
9	167	4.4	1596	4	US-08-978-277A-4
10	156.5	4.1	1503	3	US-08-976-255-14
11	148	3.9	1142	3	US-08-993-118-7
12	148	3.9	1142	3	US-08-845-528C-7
13	148	3.9	1142	4	US-09-066-281B-7
14	148	3.9	1142	4	US-09-468-433C-7
15	141	3.7	993	3	US-09-060-410-4
16	141	3.7	993	4	US-09-723-458-4
17	139	3.7	823	4	US-09-248-796A-16699
18	139	3.7	1346	2	US-08-635-121-2
19	139	3.7	1346	4	US-08-978-277A-2
20	138	3.6	1142	3	US-09-061-709-2
21	138	3.6	1142	4	US-09-899-651-2
22	138	3.6	1142	4	US-09-392-714-26
23	138	3.6	1142	4	US-09-270-437D-2
24	137	3.6	901	4	US-09-248-796A-14747
25	135.5	3.6	581	4	US-09-248-805-27
26	135	3.5	414	4	US-09-248-796A-19046
27	134.5	3.5	773	4	US-09-270-767-33141

Sequence 48359, A
Sequence 10366, A
Sequence 72, Appl
Sequence 71, Appl
Sequence 726, Appl
Sequence 1135, Ap
Sequence 10237, A
Sequence 6492, Ap
Sequence 10364, A
Sequence 2, Appl
Sequence 2, Appl
Sequence 8, Appl
Sequence 7, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 5249, Ap
Sequence 5249, Ap
Sequence 7659, Ap
Sequence 7660, Ap
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 154, App
Sequence 20328, A
Sequence 4, Appl
Sequence 5, Appl
Sequence 20227, A
Sequence 36557, A
Sequence 51774, A
Sequence 9773, Ap
Sequence 2, Appl
Sequence 944, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 14284, A
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 13, Appl
Sequence 1246, Ap
Sequence 30, Appl
Sequence 30, Appl
Sequence 1114, Ap
Sequence 30, Appl
Sequence 19073, A
Sequence 67, Appl
Sequence 67, Appl
Patent No. 5231168
Patent No. 5231168
Sequence 7, Appl
Sequence 7, Appl
Sequence 49, Appl
Sequence 2126, A
Sequence 39, Appl
Sequence 7281, Ap
Sequence 98, Appl
Sequence 22, Appl
Sequence 21, Appl
Sequence 10933, A
Sequence 6978, Ap
Sequence 7, Appl
Sequence 8902, Ap
Sequence 8903, Ap
Sequence 21836, A
Sequence 5, Appl
Sequence 3, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-893-852A-1
; Sequence 1, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLR3DT01
; CLONE: 508302
US-08-893-852A-1
; Query Match 6.6%; Score 252.5; DB 3; Length 674;
; Best Local Similarity 21.9%; Pred. No. 9.1e-13;
; Matches 164; Conservative 80; Mismatches 262; Indels 243; Gaps 32;
QY 70 LLAPLGLQKVLINSLQGLFGMFPFR--WLDAGVVSAL--RAIKGR-----112
DB 21 LLSPVMSLLSRA--WSRL-RGLGLPLEPWLVEAVKGAALVEAGLEARTPLAIPHTPWGR 77
QY 113 -----EKPAAPTAQKSLSSQLSDSPSVTSPLDWLEEGIHWOYSPDPDLKLELKAGSA 166
DB 78 RPEEAEDSGGDEDTLGLKTSLSLPEAWGLLD-DDDGMVGEREATVSP--RGQGSQ 133
QY 167 LDPAQAFLBQQLMGVBLPSSGLQSLRYNSRELSSGSPGLNQRIDNFVSVYLNPS 226
DB 134 FADGORAPL-----SPSLIIRTLQSDKNPGEEKAEGBEVAEEGVNKFSS-----YPPS 182
QY 227 YLDCPRLEVSQNSDGNSEVVGFTLTPESCLREDDH-CHPQPLSAELIPASWGGCPPL 285
DB 183 HRECCPAVE--EEDD-----EAVKKEAHTSTALSPPGSKFSTWVWSCP-- 224
QY 286 STEGLPEIHLMKRLFLQKANKQDLFTPDQDNGYHSLSEEHSLLRMDPKHCRDNPQTQ 345

Db 225 ---GEENQATEDKRTERSKGARK-----TSVSPRSGSDPRS 259
QY 346 FVPAAGDIPGNTQESTEEKIELLTTEVPLALEBESPEGCPSSSEI PNEKEFGEGRISVVD 405
Db 260 WEYRSG-----EASEKEE-----KAHEETGKGEAAPG-----287
QY 406 YSVLEGLDLPISARPACSNKLIDY-----ILGAGSSDLETSSDP-----443
Db 288 -----POSSAPAQRPQLKSWWCQPSDEESEVKALGAEKDGEAECPCIPPPSAFL 339
QY 444 -----EGEDWDEEAEDGDFSLSLSDSLDLEQDPEG-----LHLWNSFCSDVPY 487
Db 340 KAWYWPGEDETEE-EDDEEDSDSGSDEEAGEAEASSTPATGVFLKSW-----VY 391
QY 488 NPQNFTATIQTAARIVPEEPSDEKDLGSKSLDLENSQSGSLP-----ET 532
Db 392 QCGEDT-----EEDEEDSDTGSADREAREATSASTPPASAFWKAWYVRPGEDT 440
QY 533 PEH-----SSGBEDDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCGR 586
Db 441 EEEEDVDSDKEDDSEALGEAES-----DHPH-----SHPDQSAHFRGWGR 485
QY 587 DSKTPSSIVA-----ISECHTLLSKVQLLGSQ-----ESECPDSVQDVLVSGGRHTH 635
Db 486 PGKETEEEAEDWGEAEPCCFRVA--IYVPGKPPPPWAPPRLPLRLQRLKRPETPTH 543
QY 636 -----VKRKKYTFLEEVTEYYIS-----GDEDRKGPWEFEGFARDGCRFOKRIQETEDAI 683
Db 544 DDPETPLKARKYRFSEKVTVHFLAVWAGPAARQGWELQARDRFRFARRITQAEEL 603
QY 684 GYCLTFEHRERMFNRLOQTCFKGLNVLKQ 712
Db 604 SPCLTPAARARAWARLENPPALIPALTO 632

RESULT 2

US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
 LENGTH: 657 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 53041
 US-08-893-852A-3

Query Match 6.2%; Score 237; DB 3; Length 657;
 Best Local Similarity 21.8%; Pred. No. 1.9e-11;
 Matches 169; Conservative 81; Mismatches 233; Indels 292; Gaps 35;
 QY 53 SAQPTETVSVYTKLLS-QLLAPLGLKLVLSQFGMPETRLWDFAGVYSALRALKG 111
 DB 4 SPRFQ-HVLRWRDHNHFLYLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
 QY 112 REKPAAPTAQKSLSLQLDSSDPVTSPLDWLEEGIHQWYSPDLKLELKAKGSALDPAA 171
 DB 47 -----AKTVTGADQIE----- 58
 QY 172 QAFLEQOLWGVLLPSLQRLVSNRELSSPGPLNIQRIDNFSVYLLNPSYLDGCF 231
 DB 59 AALLTPTPVSG-NLLP-----HGETEESGPEQSQAQRL-----CL 94
 QY 232 PRLVSVYQNSDGNSEVGVFQTLTPESSCLREDHCHPQLSAELIPASWQGCPLSTEGLP 291
 DB 95 VEAESSPPETWGLNVDEYNA-KGQDRLREKEMERTAGKATLQAPAGLQAD----- 145
 QY 292 EIHLRLMKRLEFLOQANKG---QDLPTDQNGYHSLSEHSLLR-----MDPKHCR 340
 DB 146 -----KRLGEVAREEGVAEPAYPTSQLEGGPAENEEDGETVKTYQASAASTAPGVKP 198
 QY 341 DNPQTFVPAAGDIPGNTQESTTEKIELTTTVPALAEESSEGCPS--EIPMEKEPGE 398
 DB 199 STPVVFLGEEA-----EHQATEEK-----GTE-NKADPSNPSGSHRAWYYSREKPKQ 247
 QY 399 GRISVDVSYLEGDLPI SARPA-----CSNKLID-----YILGAS 434
 DB 248 -----EGEAKVFAHRAGQGHPCRNABEEGGPETTFVCTGNFLKAWYRPGED 296
 QY 435 SDLETSSDPE-----GEDWDEAEEDGDFSDSLSDSL 468
 DB 297 TEEDNSDSAEEDTAQTGATPHTSAFLKAWYRPGEDTEEB-----DSDSDSAEDT 350
 QY 469 EQDPEGLHWNPSVCSVDYNPQNT-----ATIQTAARI-----VPEEP 507
 DB 351 AQTGATPHT-SAFKAWYRPGEDTEENSDLSAEEDTAQTGATPHTSAFLKAWYRPG 409
 QY 508 SDSEKOLSGKSDL-----ENSSQSGSLPET-----PEHSSGREDDWES--- 545
 DB 410 EDTTEE---NSDLSAEEDTAQTGATPHTSAFLKAWYRPGEDTEDTEEDSEENVAPG 466
 QY 546 ---SADAEESLKLWNSFC-----NSDDPYNPLNFAPQTSGENEKGCRDSTPSES 594
 DB 467 DSETADSSQCLQPCRLPCKEYTKGRGEP-PL-FQVAYFLPGEK-----PES 513
 QY 595 IVAISECHTLSCVKQLIGS-----QSECPDSVORDVLGGHRTHVKKVTPLEEVTEY 650
 DB 514 PWAAPKPLRLQRLRLFPKAPTRQDPEIP-----LKARKVHFAEKVTVH 558
 QY 651 YIS-----GDEDRKGPWEAEADRCRFOKRIQETEDAIGYCLTPEHREEMENRLQ 700
 DB 559 FLAVWAGPAARRGWPEQAFDRSRFARRIAQAEEKLGLPYLTPDSRARAWRLR 613

RESULT 3
 US-08-821-818-3
 ; Sequence 3, Application US/08821818
 ; Patent No. 6146877
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Paul B.

TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
 TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/821,818
 FILING DATE:
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Chan, Albert Wai-Kit
 REGISTRATION NUMBER: 36,479
 REFERENCE/DOCKET NUMBER: 51523
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 657 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-821-818-3

Query Match 6.2%; Score 236; DB 3; Length 657;
 Best Local Similarity 21.8%; Pred. No. 2.3e-11;
 Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;
 QY 70 LLAPLGLKLVLSQFGMPETRLWDFAGVYSALRALKREKPAAPTAQKSLSLQL 129
 DB 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
 QY 130 DSDSPSVTSPLDWLEEGIHQWYSPDLKLELKAKGSALDPAAQAFLEQOLWGVLLPSS 189
 DB 47 -----AKTVTGADQIE-----AAALLTPTPVSG-NLLP-- 73
 QY 190 LQRLVSNRELSSPGPLNIQRIDNFSVYLLNPSYLDGCFRLEVSQNSDGNSEVVG 249
 DB 74 -----HGETEESGPEQSQAQRL-----CLVEAESPPETWGLSNVDE 112
 QY 250 FQTLTPESSCLREDHCHPQLSAELIPASWQGCPLSTEGLPRIHLRLMKRLEFLOQANK 309
 DB 113 YNA-KFGQDRLREKEMERTAGKATLQAPAGLQAD-----KRLGEVAREE 156
 QY 310 G---QDLPTDQNGYHSLSEHSLLR-----MDPKHCRDNPQTFVPAAGDIPGNTQ 358
 DB 157 GVAEPAYPTSQLEGGPAENEEDGETVKTYQASAASTAPGVKPSTPVVFLGEA-----EH 210
 QY 359 ESTEEKIELTTTVPALAEESSEGCPS--EIPMEKEPGEGRISVVDVSYLEGDLPI 416
 DB 211 QATEEK-----GTE-NKADPSNPSGSHRAWYYSREKPKQ-----EGEAKVE 254
 QY 417 ARPA-----CSNKLID-----YILGASDLETSSDPE----- 444
 DB 255 AHRAGQGHPCRNABEEGGPETTFVCTGNFLKAWYRPGEDTEEDNSDLSAEEDTAQ 314
 QY 445 -----GEDWDEAEEDGDFSDSLSDSLDQDPEGLHWNPSFCSDVP 486
 DB 315 TGATPHTSAFLKAWYRPGEDTEEB-----DSDSDSAEDTAQTGATPHT-SAFKAWV 367
 QY 487 YNPQNT-----ATIQTAARI-----VPEEPSDEKOLSGKSDL-----E 521

```
Db 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWVYRPGEDTEEE---NSDLSDABE 424
Qy 522 NSSQSGSLPET-----PEHSSGEEDDWS-----SADAEESKLWNSFC 560
Db 425 DTAGTATGATPHTSPFLKAWVYRPGEDTEEDTEEESENAPGDSSETADSSQCLQPQRC 484
Qy 561 -----NSDDPNPLNFKAPFOTSGENEGKCRDSTPSESIVAISECHTLLSCKVQLL 612
Db 485 LPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PESPWAAPKLPLRLQRLRLF 531
Qy 613 GS-----QESCEPDSVQRDLVSGGRHVVHKKVTFLEEVTEYYIS-----GDERKGPWE 663
Db 532 KAPTRDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPWE 576
Qy 664 EFARDGCRFOKRIOTEDAIGYCLTFEHRERMFNRLQ 700
Db 577 QFARDRSFARRIAQAEEKLGPYLTPDSRARAWRLR 613

RESULT 4
US-09-052-753B-3
; Sequence 3, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; CURRENT FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: rat
US-09-052-753B-3

Query Match 6.2%; Score 236; DB 4; Length 657;
Best Local Similarity 21.8%; Pred. No. 2.3e-11;
Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLPGLLQKVLINSOLFPGMFPTRWLDVAGVYSALRALKGRKPAAPTAQKSSLSLQ 129
Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
Qy 130 DSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAGKSGALDPAQAFLLEQQLGWVLLPSS 189
Db 47 -----AKTVTGADQIE-----AALLTTPVSG-NLLP-- 73
Qy 190 LQSRLYSNRELGGSPSPGLNTQRIQDNFVSVYLLNPSYLDGPPRLVSYQNSDGNSEYVG 249
Db 74 -----HGTEESGSPSQQAQRL-----CLVAEESPPETWGLSNVDE 112
Qy 250 FQTLTPESSCLREDHCHQPOPSAEILPASWQCPCPLSTEGLEPIHILRMKLEFLQOANK 309
Db 113 YNA-KPGQDRLREKEMERTAGKATLQAPAGQAD-----KRLGEVVAREE 156
Qy 310 G---QDLPTPDQDNGYHLSLEHSLR-----MDPKKCRDNPOTQVPAAGDIPGNTQ 358
Db 157 GVAEPAYTISQLEGGPAENEDEGETVKYQASAASIAFGYKPSIPVPLGEA-----EH 210
Qy 359 ESTEKIBLLTTEVPLALEEESPSCGPSS--EIPMEKEPGEGRISVVVDYSYLEGDLPI 416
Db 211 QATEK-----GTE-NKADPSPSPSGSHRAWYYSREKPKQ-----EGEAKVE 254
Qy 417 ARPA-----CSNKLID-----YILGGASSDLETSSDPE-----444
Db 255 AHRAGQGHPCRNAEAEEGGPETTFVCTGNAFILKAWVYRPGEDTEEDNSDSDAEEDTAQ 314
```

```
Qy 445 -----GEDWDEAEDDGFDDSSLSLSDLEQDPREGLHLWNSFCSDVP 486
Db 315 TGATPHTSAFLKAWVYRPGEDTEEE-----DSDSDAEEDTAQTGATPHT-SAFKAWV 367
Qy 487 YNPQNFT-----ATIQTAARI---VPPEPSSEKDLGSKSOL-----E 521
Db 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWVYRPGEDTEEE---NSDLSDABE 424
Qy 522 NSSQSGSLPET-----PEHSSGEEDDWS-----SADAEESKLWNSFC 560
Db 425 DTAGTATGATPHTSPFLKAWVYRPGEDTEEDTEEESENAPGDSSETADSSQCLQPQRC 484
Qy 561 -----NSDDPNPLNFKAPFOTSGENEGKCRDSTPSESIVAISECHTLLSCKVQLL 612
Db 485 LPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PESPWAAPKLPLRLQRLRLF 531
Qy 613 GS-----QESCEPDSVQRDLVSGGRHVVHKKVTFLEEVTEYYIS-----GDERKGPWE 663
Db 532 KAPTRDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARRGPWE 576
Qy 664 EFARDGCRFOKRIOTEDAIGYCLTFEHRERMFNRLQ 700
Db 577 QFARDRSFARRIAQAEEKLGPYLTPDSRARAWRLR 613

RESULT 5
US-09-052-753B-7
; Sequence 7, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; CURRENT FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 578
; TYPE: PRT
; ORGANISM: human
US-09-052-753B-7

Query Match 5.8%; Score 219; DB 4; Length 578;
Best Local Similarity 21.7%; Pred. No. 5.5e-10;
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;
```

```
Qy 53 SAQPETRVSVTWTKLLS-QLLAPLPGLLQKVLINSOLFPGMFPTRWL--DFAGV----- 102
Db 4 SPRPE-HVLHWKEAHSFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIQADA 60
Qy 103 -----YSALRALKGRKPA-APTAOKSSLSLQDSSDPSVTSPLDWLEEGIHWOY 151
Db 61 LLTPPPVSENHLPURETEGNGTPEWSKAAQRLCLDVEAQSSPKT-----WGL 108
Qy 152 SPDDLKLELKAGKSGALDPAQAFLLEQQLGWVLLPSSLQSRLYSNRELGGSPSPGLNIQ 211
Db 109 SDID---EHNGK-----PGQDGLREQEVEHTAGLP-TLQ-----PLHLQ 143
Qy 212 RIDNFSVVSYLLNPSYLDGPPRLVSYQNSDGNSEYVGQTLTPESSCLREDHCHQPOPS 271
Db 144 GADK-----KVGEVVAAREGVSALA----- 163
Qy 272 AELIPAS-WQCP---PLSTEGLEPIHILRMKLEFLQOANKGQDLPTPDQDNGYHSL 327
Db 164 ---YPTSHWEGGPAEEDTETVKKAH-----QASAASIAP-----GY----- 198
```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-821-818-2

Query Match 5.3%; Score 201; DB 3; Length 590;
Best Local Similarity 23.9%; Pred. No. 2e-08;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;

QY 245 SEVVGFTLTTPSSCLREDHCHPQLSDEL-----IPASWQGCPLSTEGPLPEIHLRM 298
DB 23 SPLMGF--LSRAWSRLRVPEA-PEWPAETVTGADQIEADAHAPPL---VPE-NHPPQ 74
QY 299 KRLEFLQQAANKGQDLPTPDQNGYHSLSEEHSLRLMDPKHCRDNP-----TQFVP--- 348
DB 75 GEAEESGTPPEGKAAQGCPLDVQANSSPPETLGLSDDDKQGDGREGQRAHTAGLPILL 134
QY 349 -----AAGDIPONTQESTEEKIELTTTEVPIALAEESPEGCPSSE----- 389
DB 135 SPGLQSAKSLGEVWAGEGVTE-----LAY-PTSHWEGCPSEEBEDGETVKKA 182
QY 390 -----IPMEKE-----PGEGRISVVDYSLGDLPIASAPACSNKLDIVLGASSDL 437
DB 183 FRASADSPGHKSTSVYCFGEAEHQATEKQENKADPPSPS-----GSHRAW 232
QY 438 ETSSDPEGEDWDE-----EADDDGFDSDSLSDSLEQDPEGLHLWNSFCV 484
DB 233 EYCSKQEGADPEPHRAGKYQLCQNAEAESEAEKVSLSVSS-----GNAFLKA 282
QY 485 DPYNPNQFT-----ATQTAARIVPEEP-----SDSEKDLGSKSDLE 521
DB 283 WYVRPGEDTDDDDSWGSAEEGKALSPTSPEHDFLKAWYVRPGEDTDD--DDSDWG 340
QY 522 NSSQSG---SLPETPEHS-----SGE--EDDWESSADEAES-----L 553
DB 341 SAESEKALSSPTSPEHDFLKAWYVRPGEDTDDQDSDWGSAEKDGLAQTFATPHTSAFL 400
QY 554 KLWNSFCNSDDPN-----PLNFKA--PFQTSGENEKGCRDSKTPSESIIVASEC-HT 603
DB 401 KTW-VCCPGEDTDDDCVWVVPEDSEADPDKSPSHEAQGC---LPGEQTEGLVEAEHS 455
QY 604 LLSCVKQLGSGES-----ECPDSVQRDVL-----SGGRHHTVKKKVTLEEVTEY 650
DB 456 LFQVAYFLGKGPAPWTPAKPLRLQRLTLRLPTQDDPETPLARKVHFSNVTVH 515
QY 651 YIS-----GDEDRKGPWBEFARDGCRFKQRIQETEDAICYLTFEHRERMFNRL 699
DB 516 FLAVWAGPAQAARRGPEWQLADRSFARRIAQAEEKLGPLYLTPAFRARAARWL 569

RESULT 8
US-09-052-753B-2
; Sequence 2, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793

US-08-821-818-2

Query Match 5.3%; Score 201; DB 3; Length 590;
Best Local Similarity 23.9%; Pred. No. 2e-08;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;

QY 245 SEVVGFTLTTPSSCLREDHCHPQLSDEL-----IPASWQGCPLSTEGPLPEIHLRM 298
DB 23 SPLMGF--LSRAWSRLRVPEA-PEWPAETVTGADQIEADAHAPPL---VPE-NHPPQ 74
QY 299 KRLEFLQQAANKGQDLPTPDQNGYHSLSEEHSLRLMDPKHCRDNP-----TQFVP--- 348
DB 75 GEAEESGTPPEGKAAQGCPLDVQANSSPPETLGLSDDDKQGDGREGQRAHTAGLPILL 134
QY 349 -----AAGDIPONTQESTEEKIELTTTEVPIALAEESPEGCPSSE----- 389
DB 135 SPGLQSAKSLGEVWAGEGVTE-----LAY-PTSHWEGCPSEEBEDGETVKKA 182
QY 390 -----IPMEKE-----PGEGRISVVDYSLGDLPIASAPACSNKLDIVLGASSDL 437
DB 183 FRASADSPGHKSTSVYCFGEAEHQATEKQENKADPPSPS-----GSHRAW 232
QY 438 ETSSDPEGEDWDE-----EADDDGFDSDSLSDSLEQDPEGLHLWNSFCV 484
DB 233 EYCSKQEGADPEPHRAGKYQLCQNAEAESEAEKVSLSVSS-----GNAFLKA 282
QY 485 DPYNPNQFT-----ATQTAARIVPEEP-----SDSEKDLGSKSDLE 521
DB 283 WYVRPGEDTDDDDSWGSAEEGKALSPTSPEHDFLKAWYVRPGEDTDD--DDSDWG 340
QY 522 NSSQSG---SLPETPEHS-----SGE--EDDWESSADEAES-----L 553
DB 341 SAESEKALSSPTSPEHDFLKAWYVRPGEDTDDQDSDWGSAEKDGLAQTFATPHTSAFL 400
QY 554 KLWNSFCNSDDPN-----PLNFKA--PFQTSGENEKGCRDSKTPSESIIVASEC-HT 603
DB 401 KTW-VCCPGEDTDDDCVWVVPEDSEADPDKSPSHEAQGC---LPGEQTEGLVEAEHS 455
QY 604 LLSCVKQLGSGES-----ECPDSVQRDVL-----SGGRHHTVKKKVTLEEVTEY 650
DB 456 LFQVAYFLGKGPAPWTPAKPLRLQRLTLRLPTQDDPETPLARKVHFSNVTVH 515
QY 651 YIS-----GDEDRKGPWBEFARDGCRFKQRIQETEDAICYLTFEHRERMFNRL 699
DB 516 FLAVWAGPAQAARRGPEWQLADRSFARRIAQAEEKLGPLYLTPAFRARAARWL 569

RESULT 9
US-08-978-277A-4
; Sequence 4, Application US/08978277A
; Patent No. 6582956
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

Db	1162	SQDSCATAAARQS	-----QVTEERAAATAQKEPSTLPNNVPAQE	120
QY	510	SEKDLGSKDLNLSQ	---SGSLPETPEHSSGEED--DW---ESSADEABSLKLWNFCN	561
Db	1201	EHGEPEGRDVLPTQOELTAAAVPLAKTEVQGEVDWLDGKVKEEQE	-----	1250
QY	562	SDDPNPLNFYKAPFQTSGENEKGCGRDSTKPSSEIVASECHTLLSKVOLLGSOESEC	PD 621	
Db	1251	-----VFVHSGPNSQKAADVITYDSE--YMGVAGQOEKSTVQSLSEEGEMET	1297	
QY	622	SVQDVLGSGRHRTHVKKKVTFLBEVTEYYIGSDEDRKGPWEE	664	
Db	1298	DVE-----KEKRETKPEQVSE---EGQETAAPHE	1325	
<p>RESULT 10</p> <p>US-08-976-255-14</p> <p>; Sequence 14, Application US/08976255</p> <p>; Patent No. 6136581</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Jono, Keith E.</p> <p>; APPLICANT: Jono, Keith E.</p> <p>; TITLE OF INVENTION: KINASE GENES AND USES</p> <p>; NUMBER OF SEQUENCES: 53</p> <p>; CORRESPONDENCE ADDRESS:</p> <p>; ADDRESSEE: Lyon & Lyon</p> <p>; STREET: 633 West Fifth Street</p> <p>; CITY: Los Angeles</p> <p>; STATE: California</p> <p>; COUNTRY: U.S.A.</p> <p>; ZIP: 90071-2066</p> <p>; COMPUTER READABLE FORM:</p> <p>; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb</p> <p>; MEDIUM TYPE: storage</p> <p>; COMPUTER: IBM Compatible</p> <p>; OPERATING SYSTEM: IBM P.C. DOS 5.0</p> <p>; SOFTWARE: FastSeq for Windows 2.0</p> <p>; CURRENT APPLICATION DATA:</p> <p>; APPLICATION NUMBER: US/08/976,255</p> <p>; FILING DATE: No. 6136581ember 21, 1997</p> <p>; CLASSIFICATION: 435</p> <p>; PRIOR APPLICATION DATA:</p> <p>; APPLICATION NUMBER: 60/031,675</p> <p>; FILING DATE: No. 6136581ember 22, 1996</p> <p>; ATTORNEY/AGENT INFORMATION:</p> <p>; NAME: Warburg, Richard J.</p> <p>; REGISTRATION NUMBER: 32,327</p> <p>; REFERENCE/DOCKET NUMBER: 229/182</p> <p>; TELECOMMUNICATION INFORMATION:</p> <p>; TELEPHONE: (213) 489-1600</p> <p>; TELEFAX: (213) 955-0440</p> <p>; TELEX: 67-3510</p> <p>; INFORMATION FOR SEQ ID NO: 14:</p> <p>; SEQUENCE CHARACTERISTICS:</p> <p>; LENGTH: 1503 amino acids</p> <p>; TYPE: amino acid</p> <p>; STRANDEDNESS: single</p> <p>; TOPOLOGY: linear</p> <p>; MOLECULE TYPE: Protein</p> <p>US-08-976-255-14</p>				
<p>Query Match 4.1%; Score 156.5; DB 3; Length 1503;</p> <p>Best Local Similarity 20.4%; Pred. No. 0.00061;</p> <p>Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;</p>				
QY	26	PRRSQAGSKFPTPLGPNSEGNMPTLLSSAQPETRVSVYWKLLSQLLAPLPGLLQKVLWS	85	
Db	826	PRRVP-----PDSLPTQGTQPTCLDIVPVE-----DCLHQDISP-----DAVTVP	867	
QY	86	QLFGGMPTRMLDPFAGVYSALRKAGREKPAAPTAQKSLQSLQSDSPSVTSLDMLWE	145	
Db	868	ETLSTDAETHSLD-----NRQDSPG-----ESETRLTESDSVLAD--DILAS	910	

QY 146 GIHWQYSPDLKELKAKGSSALDPAQAFLLEQQLMGVLLPSSLSRLYNNRELGSPPS 205
Db 911 RVSGSSUPELQGEHLNKPFFSEDDHSHR-RLKKNLEAVETLNO-----LNSKDAAKEA 962
QY 206 GPLNTQIDNFVSVYLLNPVSLDCFPRLVSVYQNSDGNSEV-----247
Db 963 GLVSALSDDSTSDQS-LLEDLSLAFPPASEPSLETTPDSLESVDVHEALLDSLGSHTPQKL 1021
QY 248 -----VGQTLTPESSCLREDCHPOPL-SABLIPAS-----WQCQPP-----LSTE 288
Db 1022 VPPDKPADSGYETENLESP---EWLHPAPEGTADSEPAATGDCGHSGLPNPNVIVISDA 1078
QY 289 GLPEIHLHLMKLEFLQANKGODLTPD-----QDNGVHSLSEEHSLLRMDPKH 338
Db 1079 G-----DGHKGTVE-TPETTAGSQGSYRDSAVFS-----1107
QY 339 CRDNPOTQVPAAGDIPGNTOESTEBKIELLTTEVPL---ALEEESPS---EGCPSSEIPME 393
Db 1108 ---DNDSEPEKSEEPVG-----TSPSALVLVQEQPLPEPVLPEQSPAAQDSCLEAR---K 1157
QY 394 KEPGGRISVDVSVYLEGDLPIARPACSNKLIIDYILGASDLETSDPDEGDW-----448
Db 1158 SQPDESCLSALHNS---SDLELRATPEPAQT-----GVPPQVHPTDEASPSVILNA 1207
QY 449 -----DREARED-----GFDSDSL-----DSDLQDPREGLHLMNSFCSDVPYN 488
Db 1208 ELSSGDDPETOQDRCTLASTGTNELLATYNMALKSLSHSEG-----1253
QY 489 PONTATTIQTAAARIVPEPSPDSEKDLGSKDLENSSQSGSLFETPEHSGBEDWESSAD 548
Db 1254 -----PKLEPDIEGKY-LGKLGVSGLDLSGDMGDADEED-ENSDD 1293
QY 549 EASLKLWNSFCNSDDPNPLNFKAPPQTSKENEKGRDSTPSESIVASECHTLLSCK 608
Db 1294 SDEDLRAFNLSLSESEDETEHPVPIILSNEDGRHLR-----1331
QY 609 VQLGQSECPDSVORDVLSGRHTRVKKKATVLEBTVFYISGDEDRKGPWEPEARD 668
Db 1332 -SLLKPTAANAPDLPED-----WKKEKAVTFDDVTYILP-----DQETPTKELGPC 1379
QY 669 G 669
Db 1380 G 1380

RESULT 11

US-08-993-118-7
; Sequence 7, Application US/08993118
; Patent No. 5997872
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,118
; FILING DATE:

CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1142
; TYPE: amino acids
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; US-08-993-118-7

Query Match 3.9%; Score 148; DB 2; Length 1142;

Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

QY 26 PRSQA---GSSKFPPTPLGPNNGNPTLLS---SAQPETRVSYMTKLSQLL-APLPGLLQ 79
Db 288 PERTQSTFEFGAQSPLQIPVSSSSSSSTLLSILFQSSPERTQSTFEGFPQSLQIPMTSSFS 347
QY 80 KVLWSOLFQGMPTRWLDFAGVYSALRKAGREKPAAPTAQKSLSSQLDSSDPSTSP 139
Db 348 STLL--SIFQSSPESAQSTFEGP-----PQSP-----LQIPGS-PSFSST 384
QY 140 LDMLEEGHWOYSP-----PDLKLELKAKGSALDPAQAFLLEQQLMGVLLPSS 189
Db 385 LLSL-----FQSSPERTHSTFEGFPQSPLOI-----PMTSSF-----SST 419
QY 190 LQSRLYNNRELGSS-----ESGFLNIQRIIDNFVSVYLL-----NPSYLDCCPR--LE 235
Db 420 LLSLILQSSPESAQSAPEFGFPQSPLOIPVSSSFSYTLSSLFQSSPERTQSTFEGFPQSPLO 479
QY 236 VSVQNSDGNSEVVGFTLTPE-----SSCLREDHCHPO-----PLSAELIPA--277
Db 480 IPVSSSSSSSTLLSILFQSSPERTQSTFEGFPQSPLOIPQSPPEGENTHSP--QIVPSLP 537
QY 278 SWQG-----CPPLSTGLPEIHLHLMKLEFLQANKGODLPT-----PDQNGVHS 324
Db 538 EWEDLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTOESTEEKIELLTTEVPALAE--378
Db 590 LSPHY--FPQSPQGEDSMSPLYFPQSPLOGE---EFQSSILQSPVSCSSSTSPSLPQSF 644
QY 379 -----ESPSEGCPSSEIPEKEPGEGRISVDVSVYLEGDLPIARPACSNKLIIDYILGAS 434
Db 645 PSSQSPPEGVQSPSLHSPQSPPEG-----MHQSPLQS-----678
QY 435 SDLETSDPDGEDWDEAEDDGFDSDSLSDLEQDP-EG-----LHL-----WNSF 481
Db 679 ----PESAPGED-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEW--718
QY 482 CSVDYPNPQNTATTIQTAAARIVPEEPSDSKOLGKSDLENSSQSGSLPETPHSGSEED 541
Db 719 ---DSLSPHFPQ-----FPPQGEDFQSSLPVSCSSSTSLSPQSPFSPSPSP 767
QY 542 DWESSADEAESLKLWNSF-----CNSDDPYNPLN--FKAPQTSKENEKGRDSTPS 592
Db 768 EGPAQSPLQRPVSSFFSYTLASLLQSSHESFPQSPPEGPAQSPLOSP-----VSSFPS 819
QY 593 ESIVASECHTLLSCKVQLLGQSESECPDS-VORDVLSGGRHTRVKKKATVLEBTVFYISGDEDRKGPWEPEARD 647
Db 820 STSSSLQSSPVSFPSTSSSLKSPESPLQSPVLSFSSSTLSL-----PFSESSSPV 875
QY 648 TEYVISGD 655
|||

Db 876 DEYTSDD 883

RESULT 12

US-08-845-528C-7

Sequence 7, Application US/08845528C

Patent No. 6027924

GENERAL INFORMATION:

APPLICANT: LUCAS, Sophie;

APPLICANT: DE SMET, Charles;

APPLICANT: BOON-FALLEUR, Thierry

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR

TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/845,528C

FILING DATE: April 25, 1997

CLASSIFICATION: 4335

ATTORNEY/AGENT INFORMATION:

NAME: Mary Anne Schofield

REGISTRATION NUMBER: 36,669

REFERENCE/DOCKET NUMBER: LUD 5455

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1142

TYPE: amino acids

STRANDEDNESS: single-stranded

TOPOLOGY: linear

US-08-845-528C-7

Query Match 3.9%; Score 148; DB 3; Length 1142;

Best Local Similarity 20.7%; Pred. No. 0.0021;

Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

Qy 26 PRRSQA---GSSKPTPLGPENSGNPTLLS--SAQPTRVSYWTKLLSLL--APLPGLLQ 79

Db 288 PERTQSTFEFGAQPQLQIPVSSSSSSLLSLFQSSPERTQSTFGFPQSLLIQIPMTSSPS 347

Qy 80 KVLWSQLFGMFTFRMLDFAGVYSALRAKGRKPAAPTAQKLSLSLDSDSDPSVTSP 139

Db 348 STLL--SIFQSSPSAQSTFEGF-----PQSP-----LQIPGS--PSFSST 384

Qy 140 LDMLEEGHMQYSP-----PDLKLELKAGKSALDPAQAFLLEQLWGVLLPSS 189

Db 385 LLSL-----FOSSPERTHTFEGFPQSPQLQI-----PWTSSF-----SST 419

Qy 190 LQSLRYNRELGSS-----PSGPNLNIQIDNFSVVSYLL-----NPSYLDGPPR--LE 235

Db 420 LLSLQSSPSAQAFEGFPQSPQLQIPVSSSSSYLLSLFQSSPERTQSTFEGFPQSPQLQ 479

Qy 236 VSYQNSGNSVGVGFQTLTP-----SSCLREDHCHIQ-----PLSALIPA-- 277

Db 480 IPVSSSSSSSTLLSLFQSSPERTQSTFEGFPQSPQLQIPQSPPEGENTHSLP--QIVPSLP 537

Qy 278 SWQG-----CPPLSTECPLPIHLRMKLEFLQKANGQDLPT-----PDQDNGYHS 324

Db 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589

Qy 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQSTEEKIELLLTTTEVPLALEE-- 378

Db 590 LSPHY--FPQSPQGEDSMSPLYFPQSPQLQE---EFQSSLQSPVSISSSTPSSLPQSF 644

Qy 379 ----ESPSECPSPSEIPMEKEGEGRISVVVDYSYLEGDLPISARPAKSNKLDIYILGGAS 434

Db 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOQ----- 678

Qy 435 SDLETSDPEGEDWDEEAEDDGFDSLSLSDSLEQDP-EG-----LHL-----WNSF 481

Db 679 ----PSAPEGD-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEWE-- 718

Qy 482 CSVDPNPNQNTATIOTAAARIVPEEPDSSEKOLSGKSDLENSQSQSLPPTPHHSGEED 541

Db 719 ---DSLSPHLFPQ-----FPQGEDFQSSLQSPVSISSSTSLSPQSPFSPQSP 767

Qy 542 DWESSADAESLKWNSF-----CNSDDPNYNPLN--FKAPFQTSGENEKGCRDSTKPS 592

Db 768 EGPAQSPQLQRPVSSFFSYTLASLLQSSHSPQSPPEGPAQSPLOQSP-----VSFP 819

Qy 593 ESIVAISECHTLILCKVQLLGSECEPDS-VQDVLSGGRHTRVKKRVTFLEE---V 647

Db 820 STSSLSQSPVSSFPSSSTSSLSKSPSPLOQSPVIFSSSTSL-----PFSESSSPV 875

Qy 648 TEYIISGD 655

Db 876 DEYTSDD 883

RESULT 13

US-09-066-281B-7

Sequence 7, Application US/09066281B

Patent No. 6475783

GENERAL INFORMATION:

APPLICANT: LUCAS, Sophie; DE SMET, Charles; BOON-FALLEUR, Thierry

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING

TITLE OF INVENTION: FOR TUMOR REJECTION ANTIGEN PRECURSOR MAGE-C1 AND MAGE-C2

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski L.L.P.

STREET: 666 Fifth Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,281B

FILING DATE: April 24, 1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/845,528

FILING DATE: April 25, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mary Anne Schofield

REGISTRATION NUMBER: 36,669

REFERENCE/DOCKET NUMBER: LUD 5455.2 US - JEL/MAS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3100

TELEFAX: (212) 752-5958

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1142

TYPE: amino acids

STRANDEDNESS: single-stranded

TOPOLOGY: linear

US-09-066-281B-7

```
Query Match          3.9%; Score 148; DB 4; Length 1142;
Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

QY 26 PRRSQA---GSKFTPLGPNNGNPTLLS--SAQETRVSYWTKLSQLL--APLPGLLQ 79
DB 288 PERTQSTFEGAQSPLQIPVSSSSSTLLSLFQSSPERTQSTFEGPQSLQIPWTSFS 347
QY 80 KVLINSQLFGMPFTRWLDFAGVYSALRAKREKPAAPTAQKSLSSLOLSDSDPSVTSP 139
DB 348 STLL--SIFQSSPESAQSTFEGF-----PQSP-----LQIPGS--PSFSST 384
QY 140 LDMLEEGHMQYSP-----PDLKLELKAKGSALDPAQAFLLEQQLWGVLLPSS 189
DB 385 LLSL-----FQSSPERTSTFEGFPQSPLOI-----PMTSSF-----SST 419
QY 190 LQRLYNRELGSS-----PSGPLNIQRIIDNFVSVYLL-----NPSYLDCEPR--LE 235
DB 420 LLSILQSSPESAQSAFEGFPQSPLOIPVSSSFVYLLSLFQSSPERTQSTFEGFPQSPLO 479
QY 236 VSYQNSDGNSEVVGFTLTPE-----SSCLREDHCHPQ-----PLSAELIPA-- 277
DB 480 IPVSSSSSTLLSLFQSSPERTQSTFEGFPQSPLOIPQSPPEGENTHSPL--QIVPSLP 537
QY 278 SWQG-----CPPLSTEGLEPEIHLRMKRLLEFLOQANKGQDLPT-----PDQONGYHS 324
DB 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQESTEEKIELLTTEVPALAE-- 378
DB 590 LSPHY--FPQSPQGEDSMSPLFPQSPLOQE---EFQSSILQSPVSISSSTPSLPSQSF 644
QY 379 ----ESPSEGCPSSEIPEKEPCEGRISVVDYSVLEGLDPLISARPACSNKLDIYLGGAS 434
DB 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOQS-----WNSF 481
QY 435 SDLETSSDPGEWDWEAEADDGFDSDSLSDLEQDP--EG-----LHL-----PES 481
DB 679 ----PESAPEGED-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEWE-- 718
QY 482 CSVDYPNQFTATIQTAAARIVPEEPSDSEKDLSGKSDLENSQSGSLPETFHSSGEED 541
DB 719 ---DLSPLHPFQ-----FPQGEDFQSSLOQSPVSISSSTSLSPQSPFSPSP 767
QY 542 DWESADAEASLKLWNSF-----CNSDDPVNPLN--FKAPFOTSGENEXGCRDSTPS 592
DB 768 EGPAQSPLOQPVSSPFTSLASLLQSSHESQSPPEGPAQSPLOQSP-----VSSPFS 819
QY 593 ESIVASECHTLTLLCKVQLLQSGSECPDS--VQRDVLSGGRHTRVKKKVTFLBE---V 647
DB 820 STSSSLQSSPVSSPSSSTSSLSKSPSPESPLQSPVIFSSTSLSS-----PFSESSSPV 875
QY 648 TEYILSGD 655
DB 876 DEYTSSTD 883

RESULT 14
US-09-468-433C-7
Sequence 7, Application US/09468433C
Patent No. 6680191
GENERAL INFORMATION:
APPLICANT: LUCAS, Sophie; BOON-FALLEUR, Thierry
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES CODING FOR
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSORS OF MEMBERS OF THE MAGE-C AN
TITLE OF INVENTION: MAGE-B FAMILIES AND USES THEREOF
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: Fulbright & Jaworski L.L.P.
STREET: 801 Pennsylvania Avenue, NW
CITY: Washington
STATE: District of Columbia
```

```
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/468.433C
FILING DATE: December 17, 1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/066,281
FILING DATE: April 24, 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/845,528
FILING DATE: April 25, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mary Anne Schofield
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: LUD 5611 JEL/WAS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 662-0200
TELEFAX: (202) 662-4643
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1142
TYPE: amino acids
STRANDEDNESS: single-stranded
TOPOLOGY: linear
US-09-468-433C-7

Query Match          3.9%; Score 148; DB 4; Length 1142;
Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 151; Conservative 90; Mismatches 257; Indels 230; Gaps 41;

QY 26 PRRSQA---GSKFTPLGPNNGNPTLLS--SAQETRVSYWTKLSQLL--APLPGLLQ 79
DB 288 PERTQSTFEGAQSPLQIPVSSSSSTLLSLFQSSPERTQSTFEGFPQSLQIPWTSFS 347
QY 80 KVLINSQLFGMPFTRWLDFAGVYSALRAKREKPAAPTAQKSLSSLOLSDSDPSVTSP 139
DB 348 STLL--SIFQSSPESAQSTFEGF-----PQSP-----LQIPGS--PSFSST 384
QY 140 LDMLEEGHMQYSP-----PDLKLELKAKGSALDPAQAFLLEQQLWGVLLPSS 189
DB 385 LLSL-----FQSSPERTSTFEGFPQSPLOI-----PMTSSF-----SST 419
QY 190 LQRLYNRELGSS-----PSGPLNIQRIIDNFVSVYLL-----NPSYLDCEPR--LE 235
DB 420 LLSILQSSPESAQSAFEGFPQSPLOIPVSSSFVYLLSLFQSSPERTQSTFEGFPQSPLO 479
QY 236 VSYQNSDGNSEVVGFTLTPE-----SSCLREDHCHPQ-----PLSAELIPA-- 277
DB 480 IPVSSSSSTLLSLFQSSPERTQSTFEGFPQSPLOIPQSPPEGENTHSPL--QIVPSLP 537
QY 278 SWQG-----CPPLSTEGLEPEIHLRMKRLLEFLOQANKGQDLPT-----PDQONGYHS 324
DB 538 EWEDSLSPHYFPQSPQGEDSLSP-HY-----FPQSPQGEDSLSPHYFPQSPQGEDS 589
QY 325 LEEHSLLRMDPKHCRD--NPTQF--VPAAGDIPGNTQESTEEKIELLTTEVPALAE-- 378
DB 590 LSPHY--FPQSPQGEDSMSPLFPQSPLOQE---EFQSSILQSPVSISSSTPSLPSQSF 644
QY 379 ----ESPSEGCPSSEIPEKEPCEGRISVVDYSVLEGLDPLISARPACSNKLDIYLGGAS 434
DB 645 PESSQSPPEGVPQSPHSPQSPPEG-----MHSQSPLOQS-----WNSF 481
QY 435 SDLETSSDPGEWDWEAEADDGFDSDSLSDLEQDP--EG-----LHL-----PES 481
DB 679 ----PESAPEGED-----SLSPLOIPQSPLEGEDSLSLHFPQSPPEWE-- 718
QY 482 CSVDYPNQFTATIQTAAARIVPEEPSDSEKDLSGKSDLENSQSGSLPETFHSSGEED 541
DB 719 ---DLSPLHPFQ-----FPQGEDFQSSLOQSPVSISSSTSLSPQSPFSPSP 767
QY 542 DWESADAEASLKLWNSF-----CNSDDPVNPLN--FKAPFOTSGENEXGCRDSTPS 592
DB 768 EGPAQSPLOQPVSSPFTSLASLLQSSHESQSPPEGPAQSPLOQSP-----VSSPFS 819
QY 593 ESIVASECHTLTLLCKVQLLQSGSECPDS--VQRDVLSGGRHTRVKKKVTFLBE---V 647
DB 820 STSSSLQSSPVSSPSSSTSSLSKSPSPESPLQSPVIFSSTSLSS-----PFSESSSPV 875
QY 648 TEYILSGD 655
DB 876 DEYTSSTD 883
```

Db 719 ---DSLSPHPQ-----FPQGEDFQSSLPVSISSSTSLSLPQSPFPSPQSP 767
Qy 542 DWSSADEAELKWNKF-----CNSDDPYPLN--FKAPFOTSGENKGRDSTKTPS 592
Db 768 EGPAQSPQRPVSVFFSYTLASLQSHESQSPPEGAQSPLOSP-----VSSFP 819
Qy 593 ESIVAISECHTLCKVQLLGSQSECPDS-VQRDVLGGGRHTHVKKKVTFLFEE---V 647
Db 820 STSSLSQSPVSPFPSSSTSSLSKSPSPLOSPVIFSSSTLSL----PFSSESSPV 875
Qy 648 TEYIISGD 655
Db 876 DEYTSSD 883

RESULT 15
US-09-060-410-4
; Sequence 4, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchineon, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-060-410-4

Query Match 3.7%; Score 141; DB 3; Length 993;
Best Local Similarity 20.7%; Pred. No. 0.0068;
Matches 106; Conservative 68; Mismatches 164; Indels 174; Gaps 24;
Qy 283 PPL-STEGLPPIHLRMKRLFFLOANKG-----QDLPTPDQDNGYHSL 326
Db 223 PPLFNWAMSAHYHIAQNESPALSGHSEYFRNFVDSCLQKIPQDRPTSE-----VLL 276
Qy 327 BEHSLLRMDPKHCRDNPQFVPAAGIPGNQTQESTEEKIELLTTEVPLALEEESPSEGCP 386
Db 277 KHRFVLRERP-----FTVIM----DLIQRKDAVRELDNLQYRKKKILFQEAENG--P 324
Qy 387 SSEIPMEKEPGE-----GRISVVVDYSYLGDLPIARPACS--NKUIDVILGASSDLE 438

Db 325 GABAPEEEEAEFYMHRAGTLTSLSSHVSFVMSISASQSSSVNLSAD-----ASDNE 378
Qy 439 TSDDPGEDEWDEAEADDGFDSDSLSDLEQDPEGLHLMNSFCSV-----DP 486
Db 379 EEEEEEEEE-EEEEEEGEPES-----REMANMQEGEHTVTSHSSIIHRLPGSDNLYDDP 431
Qy 487 YNPQ-----NFTATIQT-----ARIVPEEPSDS--E 511
Db 432 YQEMTPGPIQPPAAPPTSTSSSSARRAYCRNRDHFATIRTASLVSRQIQEHEQDSALR 491
Qy 512 KDLGKSDLENSQSGLP-----ETPEHSSGEBDDWESSADEAESLKLWNSFCNSDD 564
Db 492 EQLSGYKRMRRQHKQLLALLESRLGEREERHSLRQLEDAQ----- 533
Qy 565 PYNPLNFKAPFQTSGE-----NEKGRCDKSTPSESIVASECHTLCKVQLLGSQ 615
Db 534 -----RAGFGTEAEKLAARRHQAIGKEARAAQAERKF-----QOHILGQQ 574
Qy 616 ESECPDSVQRDVLGGGRHTHVKKKVTFLFEEVTEYIISGDEDRKGPWEFARDCGRFQKR 675
Db 575 KKEL-----AALLEAQKTY-KLRKEQLKEEL-----QENPSTPKREKAEWLRLRKEQ 621
Qy 676 IQ--ETEDAIGYCLTFEHRERMFNRLOGTCFK 705
Db 622 LQOCQAEERAGLL-----RRQRQYFELQCRQYK 649

Search completed: September 16, 2005, 10:47:17
Job time : 23.202 sec

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:46:08 ; Search time 243.057 Seconds
(without alignments)
1187.844 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MBEPTGSRKRLGRAGFRF.....RMFNRLQGTCTGKLVNLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 404927589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*

20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*

21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3804	100.0	713	16	US-10-650-482-2
2	3627	95.3	707	9	US-09-925-299-941
3	3627	95.3	707	10	US-09-925-299-941
4	2223	58.4	698	16	US-10-650-482-4
5	813	21.4	153	15	US-10-296-115-963
6	259.5	6.8	674	16	US-10-755-889-90
7	236	6.2	657	9	US-09-052-753-3
8	236	6.2	657	15	US-10-277-603-3
9	219	5.8	578	9	US-09-052-753-7
10	219	5.8	578	15	US-10-277-603-7
11	201	5.3	590	9	US-09-052-753-2
					Sequence 2, Appli
					Sequence 4, Appli
					Sequence 13, Appli
					Sequence 57852, A
					Sequence 1, Appli
					Sequence 38864, A
					Sequence 143280, A
					Sequence 62, Appli
					Sequence 44262, A
					Sequence 54269, A
					Sequence 18, Appli
					Sequence 2, Appli
					Sequence 38863, A
					Sequence 3773, Ap
					Sequence 175, App
					Sequence 462, App
					Sequence 2173, App
					Sequence 448, App
					Sequence 4505, App
					Sequence 12, Appli
					Sequence 247745, A
					Sequence 185523, A
					Sequence 7, Appli
					Sequence 79, Appli
					Sequence 7, Appli
					Sequence 54526, A
					Sequence 195416, A
					Sequence 143282, A
					Sequence 529, App
					Sequence 202963, A
					Sequence 36, Appli
					Sequence 42484, A
					Sequence 276711, A
					Sequence 5, Appli
					Sequence 195, App
					Sequence 262, App
					Sequence 106033, A
					Sequence 4755, App
					Sequence 1, Appli
					Sequence 4, Appli
					Sequence 32497, A
					Sequence 52460, A
					Sequence 7, Appli
					Sequence 165014, A
					Sequence 363, App
					Sequence 2, Appli
					Sequence 36776, A
					Sequence 2, Appli
					Sequence 73, Appli
					Sequence 54525, A
					Sequence 18, Appli
					Sequence 162500, A
					Sequence 224762, A
					Sequence 2174, App
					Sequence 202, App
					Sequence 8157, App
					Sequence 43815, A
					Sequence 52462, A
					Sequence 16, Appli
					Sequence 7, Appli
					Sequence 27, Appli
					Sequence 27, Appli
					Sequence 11, Appli
					Sequence 240, App
					Sequence 240, App
					Sequence 1, Appli
					Sequence 67803, A
					Sequence 138582, A
					Sequence 316809, A
					Sequence 217, App
					Sequence 105693, A

```

85 134.5 3.5 1253 14 US-10-363-798-2 Sequence 2, Appli
86 134.5 3.5 1692 17 US-10-732-923-8431 Sequence 8431, Ap
87 134.5 3.5 1969 9 US-09-839-479-71 Sequence 71, Appl
88 134.5 3.5 1969 15 US-10-376-537-72 Sequence 72, Appl
89 134.5 3.5 2468 15 US-10-702-148-71 Sequence 71, Appl
90 134.5 3.5 2468 16 US-10-755-889-615 Sequence 615, App
91 134.5 3.5 2468 17 US-10-489-740-216 Sequence 216, App
92 134.5 3.5 2519 18 US-10-450-763-46995 Sequence 46995, A
93 134.5 3.5 2622 18 US-10-450-763-36777 Sequence 36777, A
94 134.5 3.5 3529 17 US-10-732-923-8327 Sequence 8327, Ap
95 134.5 3.5 3571 17 US-10-732-923-8324 Sequence 8324, Ap
96 134.5 3.5 3574 17 US-10-732-923-8326 Sequence 8326, Ap
97 134.5 3.5 3574 17 US-10-732-923-8329 Sequence 8329, Ap
98 134 3.5 812 16 US-10-437-963-117056 Sequence 117056,
99 134 3.5 1439 16 US-10-719-993-803 Sequence 803, App
100 133.5 3.5 446 16 US-10-425-115-264340 Sequence 264340,

ALIGNMENTS

RESULT 1
US-10-650-482-2
; Sequence 2, Application US/10650482
; Publication No. US20040142345A1
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; APPLICANT: Jousse, Celine
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; FILE REFERENCE: PHOSPHATASE SUBUNIT
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US 60/408,679
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-482-2

Query Match 100.0%; Score 3804; DB 16; Length 713;
Best Local Similarity 100.0%; Pred. No. 1.2e-278;
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEGTGSRKRLGPRAGFRFPWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
DB 1 MPEGTGSRKRLGPRAGFRFPWPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
QY 61 SYWTKLLSOLLAPLPGLLQKVLKLSOLFQGMPTPLWDLFAGVYSALRALKGRKPAAPTA 120
DB 61 SYWTKLLSOLLAPLPGLLQKVLKLSOLFQGMPTPLWDLFAGVYSALRALKGRKPAAPTA 120
QY 121 QKSLSLLQDSSDPSTVPLDLESGIHQWQSPDPDLKLEKAKGSALDPAQAFLLEQOL 180
DB 121 QKSLSLLQDSSDPSTVPLDLESGIHQWQSPDPDLKLEKAKGSALDPAQAFLLEQOL 180
QY 181 WGVLLPSSLOSRLYSNRELSSPSGPNLQRIQDNFVSVYLLNPSYLDLCPFRLEVSYQN 240
DB 181 WGVLLPSSLOSRLYSNRELSSPSGPNLQRIQDNFVSVYLLNPSYLDLCPFRLEVSYQN 240
QY 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQGCPLSTEGLEPIHHLRMR 300
DB 241 SDGNSEVVGFTLTPESSCLREDHCHPOPLSAELIPASWQGCPLSTEGLEPIHHLRMR 300
QY 301 LEFLQAKNGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
DB 301 LEFLQAKNGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
QY 361 TEEKIELLTTEVPLALEBESSEGCPSSEIPEWKEPGEGRISVVDVSYLEGLDLPISARPA 420

```

```

Db 361 TEEKIELLTTEVPLALEBESSEGCPSSEIPEWKEPGEGRISVVDVSYLEGLDLPISARPA 420
QY 421 CSNKLIDYIILGGASSDLETSSDEGEDWDDEAEDDGDSDSSLSDSLEODPEGLHLWNS 480
DB 421 CSNKLIDYIILGGASSDLETSSDEGEDWDDEAEDDGDSDSSLSDSLEODPEGLHLWNS 480
QY 481 FCSVDPNPNQFTATIQTAAARIIVPEEPSDEKOLSGKSDLENSSQSGSLPETPHSSGEE 540
DB 481 FCSVDPNPNQFTATIQTAAARIIVPEEPSDEKOLSGKSDLENSSQSGSLPETPHSSGEE 540
QY 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRCKSTPSESIVAISE 600
DB 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGRCKSTPSESIVAISE 600
QY 601 CHTLLSCKVOLIGSQSESCPDVSQVORDVLSGGRHVKRKKVTFLEEVTEYVYISGDEDRKG 660
DB 601 CHTLLSCKVOLIGSQSESCPDVSQVORDVLSGGRHVKRKKVTFLEEVTEYVYISGDEDRKG 660
QY 661 PWEEFARDGCRFQKRIQETEDAIGCYLTFEHRERMFNRLOQTCFKGLNVLKQC 713
DB 661 PWEEFARDGCRFQKRIQETEDAIGCYLTFEHRERMFNRLOQTCFKGLNVLKQC 713

RESULT 2
US-09-925-299-941
; Sequence 941, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 941
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-941

Query Match 95.3%; Score 3627; DB 9; Length 707;
Best Local Similarity 98.8%; Pred. No. 3e-265;
Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 23 PFFPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 82
DB 17 PSVRSRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 76
QY 83 IWQLFGGMPTPLWDLFAGVYSALRALKGRKPAAPTAQKLSLSLDSSDPSTVSPDLW 142
DB 77 IWQLFGGMPTPLWDLFAGVYSALRALKGRKPAAPTAQKLSLSLDSSDPSTVSPDLW 136
QY 143 LEEGIHWQYSPDPDLKLEKAKGSALDPAQAFLLEQOLWGVLLPSSLOSRLYSNRELGS 202
DB 137 LEEGIHWQYSPDPDLKLEKAKGSALDPAQAFLLEQOLWGVLLPSSLOSRLYSNRELGS 196
QY 203 SPGGLNIQRIQDNFVSVYLLNPSYLDLCPFRLEVSYQNSDGNSEVVGFTLTPESSCLRE 262

```


Qy 491 NFTAQTAAARI VPEEPSDSKLSGKSDLENSQSGSLP-----ETPEH 535
 Db 395 EDT-----EEEDSDTGSAAEDEREAREASTTPASAFKAWVYRPGEDTEEE 443
 Qy 536 -----SSGBEDDWSADAEASIKLWNSFCNSDDPYNPLNFKAPPTOTSGENEGKCHDSK 589
 Db 444 EDEDVDSKEDDSEALGAEAS-----DHPH-----SHPDQSAHFRGNGYRPGK 488
 Qy 590 TPESIVA-----ISECHTLLSKVQLLGSQ-----ESECPSQVQDVLSSGGRHTH--- 635
 Db 489 ETEEEAAEDWGEAECPPRVA--IYVGEKPPPPWAPPRLPLRLQRLKRPETPHDPD 546
 Qy 636 -----VKKKKVTFLEEVYIIS-----GDEBKGFWEFARDGCRFKRKRIQETEDAIYGC 686
 Db 547 PETPLKARKVRFSEKVTVHFLAVWAGPAQAARQGFWEQLARDRSRFPARRIAQAQAEELSPC 606
 Qy 687 LTPHRRERMFNLQCTCFKGLNVLKQ 712
 Db 607 LTPAARARAWARLURNPLAPALATQ 632

RESULT 7

US-09-052-753-3
 ; Sequence 3, Application US/09052753A
 ; Patent No. US20010014734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Paul B.
 ; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF
 ; FILE REFERENCE: 0575/51523-B/JPW/JSG
 ; CURRENT APPLICATION NUMBER: US/09/052,753A
 ; CURRENT FILING DATE: 1998-03-31
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 657
 ; TYPE: PRT
 ; ORGANISM: Rat
 ; US-09-052-753-3

Query Match 6.2%; Score 236; DB 9; Length 657;
 Best Local Similarity 21.8%; Pred. No. 1e-08;
 Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLGLLQKVLWISQLFGCMFPTRWLDVAGVYSALRALKGREKPAAPTAKSLSLQ 129
 Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
 Qy 130 DSSDPSVTSPDLWLEEGIHQWYSPDCLKELKAGSALDPAQAQAFLEQQLWGVLLPSS 189
 Db 47 ----AKTVTGADQIE-----AAALLTPTVSG-NLLP-- 73
 Qy 190 LQSLRYNRELGSPPSPLNIQRIDNFSVSYLLNPSYLDGCFPRLEVSYQNSDGNSEVVG 249
 Db 74 ----HGETESGSPESQQAQRL-----CLVEAESPPETWGLSNVDE 112
 Qy 250 FQTLTPESSCLRDHCHPQPLSAELIPASWQCGCPPLSTGLPETHLHLMKRLFLQOANK 309
 Db 113 YNA-KPGQDDLREKEMERTAGKATLQAPAGLQAD-----KRLGVVAREE 156
 Qy 310 G----QDLPTPDQNGYHSLBEEHSLR-----MDPKHCRDNPTQFVPAAGDIPGNTQ 358
 Db 157 GVAEPAYPTSQLGGPAENEEDGETVTKYQASAASIAPGYKPTVPVFLGEA-----EH 210
 Qy 359 ESTEEKIELLTTEVPLALEBESSEGCPS--EIPMEKPGEGRISVVVDYSYLEGDLPI 416
 Db 211 QATEEK-----GTE-NKADPSNSPSSGSHRAWYYISREKPKQ-----EGEAKVE 254
 Qy 417 ARPA-----CSNKLID-----YILGASSDLETSSDPE----- 444
 Db 255 AHRAGQHPCENAEAEAGGPTTVCCTGNAPLAKAWVYRPGEDTEEDNSDSDSAEEDTAQ 314
 Qy 445 -----GEMWDEAEDDGFSDSLSLSQLEQDPRGLHLWNSFCSDVP 486

Db 315 TGAHTSAPLAKAWVYRPGEDTEEE-----DSDSASAEEDTAQTGATPHT-SAPLKAWV 367
 Qy 487 YNPQNT-----ATIQTAARI-----VPEEPSDSKLSGKSDL-----E 521
 Db 368 YRGEDETEENSOLDSAEEDTAQTGATPHTSAPLAKAWVYRPGEDTEEE---NSDLSAEE 424
 Qy 522 NSSQSGSLPET-----PEHSGBEDDWEES-----SADAEASIKLWNSFC 560
 Db 425 DTATGATGATPHTSAPLAKAWVYRPGEDTEEDTEEDSEENAVAGDSETADSSQSPCLQPC 484
 Qy 561 -----NSDDPYNPLNFKAPQOTSGENEGKCRDSTPSEISVAISECHTLLSKVQLL 612
 Db 485 LPGEKTKGGEEDP-PL-FQVAFYLPGEK-----PESWAAAPKPLRLQRLRLP 531
 Qy 613 GS-----QESCECPDSVQDVLSSGGRHTHVKKVTFLEEVYIIS-----GDEDRKGPWE 663
 Db 532 KATRDQDPEIP-----LKARKVHFAEKVTVHFLAVWAGPAQAARGPWE 576
 Qy 664 EFARDGCRFOKRIQETEDAIYCYLTFEHRERMFNLQ 700
 Db 577 QFARDRSRFPARRIAQAQAEKLGPLYLTPDSRARAWARLR 613

RESULT 8

US-10-277-603-3
 ; Sequence 3, Application US/10277603
 ; Publication No. US20030219376A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Paul B.
 ; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: A34608-B-A
 ; CURRENT APPLICATION NUMBER: US/10/277,603
 ; CURRENT FILING DATE: 2002-10-22
 ; PRIOR APPLICATION NUMBER: 09/052,753
 ; PRIOR FILING DATE: 1998-03-31
 ; PRIOR APPLICATION NUMBER: PCT/US98/05793
 ; PRIOR FILING DATE: 1998-03-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 657
 ; TYPE: PRT
 ; ORGANISM: Rat
 ; US-10-277-603-3

Query Match 6.2%; Score 236; DB 15; Length 657;
 Best Local Similarity 21.8%; Pred. No. 1e-08;
 Matches 165; Conservative 77; Mismatches 225; Indels 290; Gaps 33;

Qy 70 LLAPLGLLQKVLWISQLFGCMFPTRWLDVAGVYSALRALKGREKPAAPTAKSLSLQ 129
 Db 21 LLSPLMGLLSRA--WSRLRGPEVPEAWL----- 46
 Qy 130 DSSDPSVTSPDLWLEEGIHQWYSPDCLKELKAGSALDPAQAQAFLEQQLWGVLLPSS 189
 Db 47 ----AKTVTGADQIE-----AAALLTPTVSG-NLLP-- 73
 Qy 190 LQSLRYNRELGSPPSPLNIQRIDNFSVSYLLNPSYLDGCFPRLEVSYQNSDGNSEVVG 249
 Db 74 ----HGETESGSPESQQAQRL-----CLVEAESPPETWGLSNVDE 112
 Qy 250 FQTLTPESSCLRDHCHPQPLSAELIPASWQCGCPPLSTGLPETHLHLMKRLFLQOANK 309
 Db 113 YNA-KPGQDDLREKEMERTAGKATLQAPAGLQAD-----KRLGVVAREE 156
 Qy 310 G----QDLPTPDQNGYHSLBEEHSLR-----MDPKHCRDNPTQFVPAAGDIPGNTQ 358
 Db 157 GVAEPAYPTSQLGGPAENEEDGETVTKYQASAASIAPGYKPTVPVFLGEA-----EH 210
 Qy 359 ESTEEKIELLTTEVPLALEBESSEGCPS--EIPMEKPGEGRISVVVDYSYLEGDLPI 416
 Db 211 QATEEK-----GTE-NKADPSNSPSSGSHRAWYYISREKPKQ-----EGEAKVE 254

```
QY 417 ARPA-----CSNKLID-----YILGASSDLETSSDPE-----444
DB 255 AHRAGQGHPCRNAEAEAGGPETTFVCTGNFLKAWYRPGEDTEEDNSDSDSAEEDTAQ 314
QY 445 -----GEWDERAEDDGFDSLSLSDLEQDPEGLHLWNSFCSDVP 486
DB 315 TGATPHTSAFLKAWYRPGEDTEEB-----DSDSAEEDTAQTGATPHT-SAFLLKAW 367
QY 487 YNPQNT-----ATTQTAARI---VPEPSSDSEKDLGSKSL---E 521
DB 368 YRPGEDTEENSDDLSDAEEDTAQTGATPHTSAFLKAWYRPGEDTEEB---NSOLDABE 424
QY 522 NSSQSGSLPET-----PEHSSGREDOWES-----SADAEASLKLWNSFC 560
DB 425 DTATGATPHTSPFLKAWYRPGEDTEDTTEEDSENVAPODSETADSSQPCLOPORC 484
QY 561 -----NSDDPNPLNFKAPFQTSGENEKGCRDSTPSSIVAISECHTLTLLCKVOLL 612
DB 485 LPGEKTKGRGEBP--PL-FQVAFYLPGEK-----PESPWAAAPKLPLRLQRLRLP 531
QY 613 GS----QSESECPDSVQDRVLSGRHTHVYKRVKVTFLFEEVTEYIIS-----GDEDRKGPWE 663
DB 532 KAPTRDQDPEIP-----LKAARKVHFAEKVTVHFLAVWAGPAQAARRGPWE 576
QY 664 EFARDGCRFKRIOETDAIGYCLTFEHRERMFNRLQ 700
DB 577 QFARDRSFARRIAQAEEKLGPYLTPDSRARAWAFLR 613
```

RESULT 9

```
US-09-052-753-7
; Sequence 7, Application US/09052753A
; Patent No. US20010014734A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Fisher, Paul B.
```

```
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF
```

```
; FILE REFERENCE: 0575/51523-B/JPW/JSG
```

```
; CURRENT APPLICATION NUMBER: US/09/052,753A
```

```
; CURRENT FILING DATE: 1998-03-31
```

```
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: Patent in ver. 2.1
```

```
; SEQ ID NO 7
```

```
; LENGTH: 578
```

```
; TYPE: PRT
```

```
; ORGANISM: Human
```

```
US-09-052-753-7
```

```
Query Match 5.8%; Score 219; DB 9; Length 578;
```

```
Best Local Similarity 21.7%; Pred. No. 1.6e-07;
```

```
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;
```

```
QY 53 SAQPETRVSYWTKLLS-QLLAPLPLGLLQKVLVWSQLFCGMFPTRWL--DFAGV-----102
DB 4 SPRPE-HVLHWKEAHSFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIQADA 60
QY 103 -----YSALRAKLGREKPA-APTAQKSLSSQLDSDSPSVTSPLDWLEEGIHQY 151
DB 61 LLTPPPVSENHPLPLETEGNGTPEWSKAAQRLCLDVEAQSSPPKT-----WGL 108
QY 152 SPDDLKLELKAGSALDPAQAFLLEQOLMGVVELLPSSLSQRLSYNRELSSPSGGLNIQ 211
DB 109 SDID---EHNGK-----PGQDGLREQVEHTAGLP-TLQ-----PLHLQ 143
QY 212 RIDNFSVSYLLNPSYLDLCPFRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLS 271
DB 144 GADK-----KVGEVVARBEGVSELA-----163
QY 272 AELIPAS-WQGCP---PLSTEGPLPEIHHLMKRLFLQQAANKGODLPTPDODNGYHSLEE 327
DB 164 ---YPTSHWGGPAEDEDTEVTKAH-----QASAASIAP-----GY-----198
QY 328 EHSLLRMDPKHCRNPOTQVFPAAAGDIPGNTQSTEEKI-----ELLTTEVPLALE 377
```

```
DB 199 -----KPSTSVYCPGEAHEHRAATEBEKGTDNKAEPSSGSHSRFWFVHTRRPKQEG 246
QY 378 EESPSECCPSSEIP--MEKEPGEGRISVVDYSYLEGDLPLISARPACSNKLIDYILGGAS 434
DB 247 ETKPEQHRAGQSHPCQNAESEEGPETS-----CS-----GS 279
QY 435 SDLETSSDPGEGDWEAEDDGFDSLSLSDLEQDPEGLHLWNSFCSDVPYNPQNFTA 494
DB 280 AFLKAWYRPGEDTEEB-EDPDLD-----AEED-----TAHTCTTPH-----TS 318
QY 495 TIQTAARIVPEEPSDSKDLGSKDLSNLSQSGSLPET-----PEHSSGREDWE 544
DB 319 AFLKAWYRPGEDTEEDDGDWDSAEEDAAQSCCTTPTHTSAFLKAWYRPGEDTEEDDSE 378
QY 545 SSAD-EAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSSIVAISECHT 603
DB 379 NVAPVDSETV-----DSCQSTQHCLFV-----BKTGCGGEAREPPFQVA-----417
QY 604 LLSCKVOLLGSOES-----ECPDSVQDRVLS-----GGRHTHVYKRVKVTFLFEEVTEY 650
DB 418 -----FVLPQKAPPWAAPKLPLRLQKRLRSFKAPARNQGPPEIPLAKRKHVFSEKVTVH 472
QY 651 YIS-----GDEDRKGPWEFARDGCRFKRIOETDAIGYCLTFEHRERMFNRLQ 700
DB 473 FLAVWAGPAQAARRGPWEQFARDRSFARRIAQAEEQLGPLYLTPAFARAWAFLR 527
```

RESULT 10

```
US-10-277-603-7
```

```
; Sequence 7, Application US/10277603
```

```
; Publication No. US20030219376A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Fisher, Paul B.
```

```
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
```

```
; TITLE OF INVENTION: THEREOF
```

```
; FILE REFERENCE: A34608-B-A
```

```
; CURRENT APPLICATION NUMBER: US/10/277,603
```

```
; CURRENT FILING DATE: 2002-10-22
```

```
; PRIOR APPLICATION NUMBER: 09/052,753
```

```
; PRIOR FILING DATE: 1998-03-31
```

```
; PRIOR APPLICATION NUMBER: PCT/US98/05793
```

```
; PRIOR FILING DATE: 1998-03-20
```

```
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 7
```

```
; LENGTH: 578
```

```
; TYPE: PRT
```

```
; ORGANISM: Human
```

```
US-10-277-603-7
```

```
Query Match 5.8%; Score 219; DB 15; Length 578;
```

```
Best Local Similarity 21.7%; Pred. No. 1.6e-07;
```

```
Matches 155; Conservative 79; Mismatches 223; Indels 258; Gaps 34;
```

```
QY 53 SAQPETRVSYWTKLLS-QLLAPLPLGLLQKVLVWSQLFCGMFPTRWL--DFAGV-----102
DB 4 SPRPE-HVLHWKEAHSFYLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIQADA 60
QY 103 -----YSALRAKLGREKPA-APTAQKSLSSQLDSDSPSVTSPLDWLEEGIHQY 151
DB 61 LLTPPPVSENHPLPLETEGNGTPEWSKAAQRLCLDVEAQSSPPKT-----WGL 108
QY 152 SPDDLKLELKAGSALDPAQAFLLEQOLMGVVELLPSSLSQRLSYNRELSSPSGGLNIQ 211
DB 109 SDID---EHNGK-----PGQDGLREQVEHTAGLP-TLQ-----PLHLQ 143
QY 212 RIDNFSVSYLLNPSYLDLCPFRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLS 271
DB 144 GADK-----KVGEVVARBEGVSELA-----163
QY 272 AELIPAS-WQGCP---PLSTEGPLPEIHHLMKRLFLQQAANKGODLPTPDODNGYHSLEE 327
```

Db 164 ---YPTSHWGGPAEDDEDTETVKKAH-----QASAAAIAP-----GY----- 198
Qy 328 EHSLLRMDPKHCRDNPQFVPAAGDIPGNTQESTEEKI-----ELLTTEVPLALE 377
Db 199 -----KPTSVYCPGEAEHRAATEKGTDNKAEPGSHSRPWEYHTRERPKQEG 246
Qy 378 BEPSECCPSEIP---MEKEPGEGRISVVDVSYLEGLDLPISARPACSNKLIIDYLGAS 434
Db 247 ETKEPQHRAGQSHPCQNAESEGGPETS-----CS-----GS 279
Qy 435 SDLETSDDPGEOWDEAEADGDDSDSLSDLEQDPEGLHLWNSFCSVDYPNPQNFTA 494
Db 280 AFLKAWYRPGEDTEEE-EDPDLS-----AEED-----TAHTCTTPH-----TS 318
Qy 495 TIQTAARIVPEEPSDSKOLSGKSDLENSQSGSLPET-----PEHSGEEDDWE 544
Db 319 AFLKAWYRPGEDTEEDDDGWDSDAEADAQSCITPHTSAFLKAWYRPGEDTEEDDSE 378
Qy 545 SSAD-EAESLKLWNSFCNSDDPYNLFKAPFQTSGENEKGCRDSTPSSSIVAISECHT 603
Db 379 NVAPVDSETV---DSQSTQHCLFV-----EKTGCGEABEPFPQVA----- 417
Qy 604 LLSCKVQLLSQES-----ECPDSVQDVL-----GGRHTHVKRKVTPLLEEVTEY 650
Db 418 -----FYLPGQKAPPWAAPKLPLRLQRLSKFAPARNQGPPELPLGRKRVHSEKVTVH 472
Qy 651 YIS-----GBEDRKGWEEFARDGCRFQKRIQETEDAIGVCLTFEHRMFMNRLQ 700
Db 473 FLAWAGPAQAARGFWEQIARDRSRFAIRIAQAEKJGPLYLTPAFARAWTLRL 527

RESULT 11
US-09-552-753-2
; Sequence 2, Application US/09052753A
; Patent No. US20010014734A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES THEREOF
; FILE REFERENCE: 0575/51523-B/JPW/JSG
; CURRENT APPLICATION NUMBER: US/09/052,753A
; CURRENT FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Rat
US-09-552-753-2

Query Match 5.3%; Score 201; DB 9; Length 590;
Best Local Similarity 23.9%; Pred. No. 3.8e-06;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;
Qy 245 SEVVGFQTLTPESSCLREDHCHPOPLSAEL-----IPASWQCCPPLSTGLPEIHLRM 298
Db 23 SPLMGF--LSRAWSLRVPEA-PEPWAETVTGADQIADAHAPPL-----VPE-NHPQ 74
Qy 299 KRLEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNP-----TQFVP--- 348
Db 75 GEABESGTPEEGKAAQPCLDVQANSPPETGLSDDDKQGGQPGQRAHTAGLPILL 134
Qy 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSGCPSS- 389
Db 135 SPGLQADKSLGEVAGEGVTE-----LAV-PTSHWEGCPSEEEDEGTVKA 182
Qy 390 -----IPMEKE-----PGEGRISVVDVSYLEGLDLPISARPACSNKLIIDYLGASDL 437
Db 183 FRASADSPGHKSSTSVYCPGEAEHQATEEKQTKENKADPPSSPS-----GSHRAW 232
Qy 438 ETSSDPGEOWDE-----EAEDDGPDSLSLSLEQDPEGLHLWNSFCSV 484
Db 233 EYCSKQGEADPEPHRAGKYQLCONAEAEAEAEKVSLSVSS-----GNAFLKA 282

Qy 485 DPYNPNFT-----ATIQTAAIRIVPEEP-----SDSEKOLSGKSDLE 521
Db 283 WYVRPGEDTDDDDSDWGSAAEEGKALSPTSSEHDFLKAWYRPGEDTDD--DDSDWG 340
Qy 522 NSSQSG---SLPPTPEHS-----SGE--EDDWESSADEAES-----L 553
Db 341 SAESEKALSSPTSSEHDFLKAWYRPGEDTDDQSDWGSAAEKDGLAQTFATPHTSAPL 400
Qy 554 KLWNSFCNSDDPFYN-----PLNFKA--PFQTSGENEKGCRDSTPSSSIVAISEC-HT 603
Db 401 KTW-VCCPGEDTDDCEVVVVPEDSEADPDKSPSHEAQGC-----LPGEQTEGLVAEHS 455
Qy 604 LLSCKVQLLSQES-----ECPDSVQDVL-----SGRTHYVKKVTPLEEVTEY 650
Db 456 LFQVAFYLFCEKPAFPWPTAPKLPLRLQRLTLRLTPTQDDPETPLURARKVHSEKVTVH 515
Qy 651 YIS-----GBEDRKGWEEFARDGCRFQKRIQETEDAIGVCLTFEHRMFMNRL 699
Db 516 FLAWAGPAQAARGFWEQIARDRSRFAIRIAQAEKJGPLYLTPAFARAWTLRL 569

RESULT 12
US-10-277-603-2
; Sequence 2, Application US/10277603
; Publication No. US20030219376A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: PROGRESSION ELEVATED GENE-3 AND USES
; FILE REFERENCE: A34608-B-A
; CURRENT APPLICATION NUMBER: US/10/277,603
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/052,753
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Rat
US-10-277-603-2

Query Match 5.3%; Score 201; DB 15; Length 590;
Best Local Similarity 23.9%; Pred. No. 3.8e-06;
Matches 142; Conservative 56; Mismatches 210; Indels 186; Gaps 29;
Qy 245 SEVVGFQTLTPESSCLREDHCHPOPLSAEL-----IPASWQCCPPLSTGLPEIHLRM 298
Db 23 SPLMGF--LSRAWSLRVPEA-PEPWAETVTGADQIADAHAPPL-----VPE-NHPQ 74
Qy 299 KRLEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNP-----TQFVP--- 348
Db 75 GEABESGTPEEGKAAQPCLDVQANSPPETGLSDDDKQGGQPGQRAHTAGLPILL 134
Qy 349 -----AAGDIPGNTQESTEEKIELLTTEVPLALEEESPSGCPSS- 389
Db 135 SPGLQADKSLGEVAGEGVTE-----LAV-PTSHWEGCPSEEEDEGTVKA 182
Qy 390 -----IPMEKE-----PGEGRISVVDVSYLEGLDLPISARPACSNKLIIDYLGASDL 437
Db 183 FRASADSPGHKSSTSVYCPGEAEHQATEEKQTKENKADPPSSPS-----GSHRAW 232
Qy 438 ETSSDPGEOWDE-----EAEDDGPDSLSLSLEQDPEGLHLWNSFCSV 484
Db 233 EYCSKQGEADPEPHRAGKYQLCONAEAEAEAEKVSLSVSS-----GNAFLKA 282
Qy 485 DPYNPNFT-----ATIQTAAIRIVPEEP-----SDSEKOLSGKSDLE 521
Db 283 WYVRPGEDTDDDDSDWGSAAEEGKALSPTSSEHDFLKAWYRPGEDTDD--DDSDWG 340
Qy 522 NSSQSG---SLPPTPEHS-----SGE--EDDWESSADEAES-----L 553

```

Db 341 SAEEBKGALSSPTSEPHDFLKAWVYRPGEDTEDQDSDWGAEXDGLAQATPATHTSAFL 400
Qy 554 KLNWSFNCSDDPYN-----PLNPKA--PQOTSGENKGCGRDSTKTPSESTIVALSEC-HT 603
Db 401 KTW-VCCPGEDTEDDCVVPVPEDEAAADPKSPSHEAQC-----LPGEQTEGLVEABHS 455
Qy 604 LLSCKVQLLGSQES-----ECPOSDVQDVL-----SGGRHTHVKKKVTLEEVTEY 650
Db 456 LFOVAFYLPGEKPPWTAAPKPLRLQRLTLRLTPTQDQDPETPLRKARKVHFSENVTVH 515
Qy 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIQCYCLTFEHRERMFNRL 699
Db 516 FLAVWAGPAQAARCGPWFQQLARDRSFARRIAQAEKLGPLYTTPAFRARAWARL 569

RESULT 13
US-09-902-432-4
; Sequence 4, Application US/09902432
; Publication No. US2002016002A1
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A 070156.0597
; CURRENT APPLICATION NUMBER: US/09/902,432
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-902-432-4

Query Match 4.4%; Score 167; DB 9; Length 1596;
Best Local Similarity 19.8%; Pred. No. 0.0061;
Matches 151; Conservative 112; Mismatches 286; Indels 214; Gaps 35;

Qy 6 GGSRKRL-----GPR-----AGFRFWPPPPRRSQAGSSKEPT 38
Db 673 GSSKRRKAKSSDDDEGGFRTLGDSHRAEASKDKEAGTDAVPASTQEQDQAQSSSPE 732
Qy 39 PLGPENSGNPTLLSSAQOETRVSYWTKLLSQLLAPLGLQKV-----LIWSQLFGGMF 92
Db 733 PAGSPSEGE-----GVSTW-ESFKLVTPRKSKSLBEKARDSSVEQLSTEIE 780
Qy 93 PTR---WLDFAGVYSALRAKREKPAAPTAKSLS-----SLQDSSDPVSPTS--PLDW 142
Db 781 PSRESWV-----SIKKFIPGRKRRADGQEQATVBDSDGFEVNEIDDPNPAVAVPLS- 833
Qy 143 LEEGHWOYSPDLEKLEKAGSALDPAQAFLEQLQWGVLLPSSLSQSLYSNRELGS 202
Db 834 -----EYNAVE-REKMEAGNTELP-----QLLGA-----VYVSEELSK 866
Qy 203 SPSPGLNQTQIDNFSVSYL--LNPSYLDCE-----PRLEVSQNSDGNSE 246
Db 867 TLVHTVSVAVIDGTNAVTSVERSPSWISASVTEPLEHTAGEAMPVEEVTEKIDIAEET 926
Qy 247 VVGFTLTPRESSCLREHCHPO-PLSABELIPASWQGCPLPSTEGPEIHHLMKRLBFLQ 305
Db 927 PVLATQTL-PEGKDAHDDMVTSEVDPTSEAVTAT-ETSEALRTEEVTEASGAE-ETTDVVS 983
Qy 306 QANKQDLP-----TPDQ--DNGVHSLREE-----HSLLRMDPKHCRDN-----PTQFVPA 349
Db 984 AVSQTDSPTTEATPQEVESGVLDTTEBERQTQAILQAVADKVKESQVPATQTVQR 1043
```

```

Qy 350 AGDIPGNTQSTBEKIELLTTEVPLALEEBSPEGCPSPSEIPMEKEPGEGRISVVDYSYL 409
Db 1044 TGSKALEKFVEVEDSEVLASEKEKDVPKGPQVEAGAEHLAQOSETGQATPESLEVPEV 1103
Qy 410 EGDLPISARPACS-----NKLIIDYILGASSDLTSTSDPEG-----EDWDEEAEDDGF 458
Db 1104 TADVHDVA--TCQVIKQLQMEQAVAPESSETLTDSETNGSTPLADSDTADGTQDQDETID 1161
Qy 459 SDSLSDSLQEDQEPGLHLWNSFCSDVPYNPQNFATIQTAARIVPREPS-----D 509
Db 1162 SQDSKATAAARQS-----QVTEBEAATAQKESEPSTLPNNVPAQE 1200
Qy 510 SEKDLSKSDLENSQ---SGSLPETPEHSGBED--DW---ESSADEAESLKLWNSFCN 561
Db 1201 EHGEFGRDVLPTQOQELTAAAVPLAKTEVGQEGEVLDLGEKVKEQE----- 1250
Qy 562 SDDPYNPLNFKAPQTSGENEKGRCKTSPSEIVALSECHTLLSKVQLLGGQESCPD 621
Db 1251 -----VFVHSGPNSQKAADVTDSE--VMGVAGCQEKESTEVQSLSEEGEMET 1297
Qy 622 SVQRDVLSSGGRHTHVKKKVTFLSEVTEYIISGDEORKGPMEE 664
Db 1298 DVE-----KEKRETKPEQVSE---EGEQETAPEHE 1325

RESULT 14
US-10-097-534-13
; Sequence 13, Application US/10097534
; Publication No. US20030049607A1
; GENERAL INFORMATION:
; APPLICANT: GREENER, TSVIKA
; APPLICANT: MOSKOWITZ, HAIM
; APPLICANT: REISS, YUVAL
; APPLICANT: ALROY, IRIS
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION OF VIRAL
; FILE REFERENCE: PLV-001.01
; CURRENT APPLICATION NUMBER: US/10/097,534
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/275,224
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/308,958
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/340,170
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 13
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-534-13

Query Match 4.2%; Score 160.5; DB 14; Length 1562;
Best Local Similarity 20.9%; Pred. No. 0.018;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;

Qy 31 AGSKKPTTP-LGPNENSGNPTLLSSAQOETRVSYWT-KLLSQLLAPLPG-LLOKVLWSQL 87
Db 123 SGALRATTPSVTVKNNSAAPFKSIGADETVQGGSRRLISFSLDSDFQAMGLKK----- 175
Qy 88 FGGMF--PTRWLDFA---GVYSALRAL--KGRKPAAPTAKSKLSLSLQDSSDPVSPTSPL 140
Db 176 --GMFNPDPYLVKISIQPKHSIFPALPHHQR-----RSKLI-----NTVNP 219
Qy 141 DWLEEGHWOYSPDP-LKLEKAGSALDPAQAAP-----LLEQQLWGVLLPSS 189
Db 220 -WQAEQSFVSLEPTDVLIEVQDKFAKSRPIIKRFLGKLSMPVQRLIERHAIGDRVVSYT 278
Qy 190 LQSLYXSNRELG-----SSPSGPNLQIRDNFVSVYLLNPSYLD 230
Db 279 LGRRLPDTHVSGQLQFRFEITSSIHDPDEISISTEPESAQIQD--SPMNLMESSGSE- 335
```



```
Qy 231 PPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWGCPPLSTEG 290
Db 336 -PREAPESSESWKPEQLGEGSVDRPCNQSIELSRPAEEAAVITEAGDQGMVSVGPEGA 394
Qy 291 PEIHLRMKLEFLQOANKG-QDLPTDQDNGYHSLEEHSLRMDPKHCRDNPQTQVPA 349
Db 395 GEL-----LAQVKDIQAPSAELAEQDLGEEASALLE- 431
Qy 350 AGDIPGNTQESTEKIELLTTEVPLALEE---ESPSEGCSPSEIPMEKEPGEGRISVVYD 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEEGEGDVST-----LEQEGELQ----- 479
Qy 407 SYLEGDLPIGARPACSNKL-IDYILGGASSDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVYKRSRP-CSLPVSELETIVIASACGDPETPRTHYIRHTLLHSMPSAQGSAE 536
Qy 453 EDGCFDSDSLSDLEQDPEGLHWNFSFCSVDYPNQNTATIQTAARIVPEPSPSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLGKSDLENSQSGLPET--PEHSSG-----BEDDWESSADEASLKLWNSF 559
Db 572 D-----REEPGATPGTAHPGSHGHPPLSLANGAQQDTHPSTGSEDS----- 616
Qy 560 CNSDDPNPLNFKAPFTSGENEKGRDSTKTPSESIVAISECHTLTLLSKVQLLGSQSEC 619
Db 617 -----SPRQGGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYASCY 660
Qy 620 PDSVQRDVLSGGR---HTHVKRKVTFLEEVTEYIISGDEDRK----- 659
Db 661 P-----SCYNGNRFASHTRFSSVDSAKISESTVFSSQDDDEEENSAPESVPDSMQSPELD 715
Qy 660 -----GPW-BEFPARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLOQTC 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGE 758

RESULT 15
US-10-450-763-57852
; Sequence 57852, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 57852
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1524)..(1556)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: EMATRIX, accession number PF00632C, p-values=3.302e-23, raw score
; OTHER INFORMATION: 20.66
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1562)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-value=4.3e-127, Pfam score of 435.6
US-10-450-763-57852
```

Query Match

4.2%; Score 160.5; DB 18; Length 1562;

```
Best Local Similarity 20.9%; Pred. No. 0.018;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;
Qy 31 AGSKKEPTP-LGPNSENGNPTLLSSAQPETRVSYWT-KLLSQLLAPLPLGL-LQKVLWISQL 87
Db 123 SGALRATTSVTWNSAAPIFKISIGADETVQGGSRRLISFSLDFQAMGLKK----- 175
Qy 88 FGGMF--PTRWLDF--GVYSALRAL--KGREKPAAPTAQKLSLSLQDSSPVSPTSPL 140
Db 176 --GMFPNPDYKISIQPGKHSIFPALPHHGQR-----RSKIIG-----NTWNP 219
Qy 141 DWLEIGHWQYSPD-LKLELKAKGALDPAQAFA-----LLEQQLWGVLLPSS 189
Db 220 -WQAEQFSFVSLPTDLEIEVDKFAKSRPIIKRFLGKLSMPVQRLERHAIGDRVVSYT 278
Qy 190 LQRLYSNRELG-----SSPSGFLNIQRIDNFSVSVLLNPSVLD 230
Db 279 LGRRLTDHVGQQLQRFPEITSIHPDDDEISLSTEPESAQIQD--SPMNNLESSEGE- 335
Qy 231 PPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLSAELIPASWGCPPLSTEG 290
Db 336 -PREAPESSESWKPEQLGEGSVDRPCNQSIELSRPAEEAAVITEAGDQGMVSVGPEGA 394
Qy 291 PEIHLRMKLEFLQOANKG-QDLPTDQDNGYHSLEEHSLRMDPKHCRDNPQTQVPA 349
Db 395 GEL-----LAQVKDIQAPSAELAEQDLGEEASALLE- 431
Qy 350 AGDIPGNTQESTEKIELLTTEVPLALEE---ESPSEGCSPSEIPMEKEPGEGRISVVYD 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEEGEGDVST-----LEQEGELQ----- 479
Qy 407 SYLEGDLPIGARPACSNKL-IDYILGGASSDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVYKRSRP-CSLPVSELETIVIASACGDPETPRTHYIRHTLLHSMPSAQGSAE 536
Qy 453 EDGCFDSDSLSDLEQDPEGLHWNFSFCSVDYPNQNTATIQTAARIVPEPSPSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLGKSDLENSQSGLPET--PEHSSG-----BEDDWESSADEASLKLWNSF 559
Db 572 D-----REEPGATPGTAHPGSHGHPPLSLANGAQQDTHPSTGSEDS----- 616
Qy 560 CNSDDPNPLNFKAPFTSGENEKGRDSTKTPSESIVAISECHTLTLLSKVQLLGSQSEC 619
Db 617 -----SPRQGGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYASCY 660
Qy 620 PDSVQRDVLSGGR---HTHVKRKVTFLEEVTEYIISGDEDRK----- 659
Db 661 P-----SCYNGNRFASHTRFSSVDSAKISESTVFSSQDDDEEENSAPESVPDSMQSPELD 715
Qy 660 -----GPW-BEFPARDGCRFQKRIQETDAIGYCLTFEHRERMFNRLOQTC 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GPSNRREGE 758
```

Search completed: September 16, 2005, 11:04:29
Job time : 245.057 secs

THIS PAGE BLANK (USPTO,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:40:09 ; Search time 182.419 Seconds
(without alignments)
4565.254 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEFGTGGRKRLGPRAGFRF.....RMFNRLQGTCTFKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6959266 seqs, 1168006243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Databases :

1: Pending Patents_AA_Main.*
2: /cgn2_6/ptodata/1/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/1/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/1/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/1/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/1/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/1/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/1/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/1/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/1/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/1/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/1/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/1/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/1/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/1/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/1/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/1/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/1/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/1/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/1/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/1/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/1/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/1/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/1/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/1/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/1/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/1/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/1/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/1/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/1/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/1/paa/US105 COMB.pcp.*
32: /cgn2_6/ptodata/1/paa/US106 COMB.pcp.*
33: /cgn2_6/ptodata/1/paa/US107 COMB.pcp.*
34: /cgn2_6/ptodata/1/paa/US108 COMB.pcp.*
35: /cgn2_6/ptodata/1/paa/US109 COMB.pcp.*
36: /cgn2_6/ptodata/1/paa/US110 COMB.pcp.*
37: /cgn2_6/ptodata/1/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3804	100.0	713	32	US-10-650-482-2	Sequence 2, Appl1
2	3799	99.9	713	27	US-10-170-205E-23916	Sequence 23916, A
3	3799	99.9	713	37	US-60-452-680-17655	Sequence 17655, A
4	3794	99.7	713	20	US-09-629-469A-18965	Sequence 18965, A
5	3794	99.7	713	35	US-10-917-503-18965	Sequence 18965, A
6	3789	99.6	720	22	US-09-757-028-2348	Sequence 2348, Ap
7	3789	99.6	720	28	US-10-222-911-2348	Sequence 2348, Ap
8	3627	95.3	707	1	PCT-US00-05883-941	Sequence 941, App
9	3627	95.3	707	24	US-09-925-299-941	Sequence 941, App
10	2223	58.4	698	32	US-10-650-482-4	Sequence 4, Appl1
11	813	21.4	153	1	PCT-US00-35017A-963	Sequence 963, App
12	813	21.4	153	28	US-10-296-115-963	Sequence 963, App
13	511	13.4	101	1	PCT-US01-14827-15743	Sequence 15743, A
14	402	10.6	73	37	US-60-177-571-4491	Sequence 4491, Ap
15	399	10.5	73	37	US-60-162-247-4518	Sequence 4518, Ap
16	399	10.5	73	37	US-60-169-840-8186	Sequence 8186, Ap
17	334	8.8	67	37	US-60-196-718-7356	Sequence 7356, Ap
18	263	6.9	697	34	US-10-821-234-905	Sequence 905, App
19	260.5	6.8	674	27	US-10-170-205E-18860	Sequence 18860, A
20	260.5	6.8	674	28	US-10-219-051B-549	Sequence 549, App
21	260.5	6.8	674	28	US-10-213-051B-553	Sequence 553, App
22	260.5	6.8	674	28	US-10-219-051B-10552	Sequence 10552, A
23	260.5	6.8	674	37	US-60-443-566-4148	Sequence 4148, Ap
24	260.5	6.8	674	37	US-60-452-680-23269	Sequence 23269, A
25	260.5	6.8	674	37	US-60-453-050-14551	Sequence 14551, A
26	260.5	6.8	674	37	US-60-453-135-14551	Sequence 14551, A
27	260.5	6.8	674	37	US-60-455-444-7838	Sequence 7838, Ap
28	260.5	6.8	674	37	US-60-465-241-7838	Sequence 7838, Ap
29	260.5	6.8	674	37	US-60-466-412-14551	Sequence 14551, A
30	259.5	6.8	674	20	US-09-629-469A-11494	Sequence 11494, A
31	259.5	6.8	674	33	US-10-755-889-90	Sequence 90, Appl
32	259.5	6.8	674	35	US-10-917-503-11494	Sequence 11494, A
33	259.5	6.8	674	37	US-60-440-068-90	Sequence 90, Appl
34	259.5	6.8	674	37	US-60-469-757-90	Sequence 90, Appl
35	252.5	6.6	674	12	US-08-893-852-1	Sequence 1, Appl1
36	237	6.2	657	12	US-08-893-852-3	Sequence 3, Appl1
37	236	6.2	657	28	US-10-274-666-3	Sequence 3, Appl1
38	236	6.2	657	28	US-10-277-603-3	Sequence 3, Appl1
39	219	5.8	578	16	US-09-236-772-2	Sequence 2, Appl1
40	219	5.8	578	28	US-10-274-666-7	Sequence 7, Appl1
41	219	5.8	578	28	US-10-277-603-7	Sequence 7, Appl1
42	218.5	5.7	627	21	US-09-724-676A-55843	Sequence 55843, A
43	218.5	5.7	627	21	US-09-724-676A-55843	Sequence 55843, A
44	201	5.3	590	12	US-08-893-852-4	Sequence 4, Appl1
45	201	5.3	590	28	US-10-274-666-2	Sequence 2, Appl1
46	201	5.3	590	28	US-10-277-603-2	Sequence 2, Appl1
47	170	4.5	1095	1	PCT-US03-38896-184	Sequence 184, App
48	170	4.5	1095	27	US-10-170-205E-22541	Sequence 22541, A
49	170	4.5	1095	37	US-60-452-680-14318	Sequence 14318, A
50	170	4.5	1095	37	US-60-453-050-8816	Sequence 8816, Ap
51	170	4.5	1095	37	US-60-453-135-8816	Sequence 8816, Ap
52	170	4.5	1095	37	US-60-466-412-8816	Sequence 8816, Ap
53	168	4.4	1454	1	PCT-US03-28227-5026	Sequence 5026, Ap
54	167	4.4	1596	1	PCT-US97-06830-4	Sequence 4, Appl1
55	167	4.4	1596	10	US-08-665-401-4	Sequence 4, Appl1
56	167	4.4	1596	24	US-09-902-432-4	Sequence 4, Appl1
57	167	4.4	1596	28	US-10-213-051B-8255	Sequence 8255, Ap
58	160.5	4.2	1562	1	PCT-US01-08631-57852	Sequence 57852, A
59	160.5	4.2	1562	26	US-10-097-534-13	Sequence 13, Appl1
60	160.5	4.2	1585	30	US-10-487-132-1	Sequence 1, Appl1
61	160.5	4.2	1607	18	US-09-488-725A-2337	Sequence 2337, Ap
62	160.5	4.2	1607	28	US-10-258-898A-2337	Sequence 2337, Ap
63	160.5	4.2	1664	18	US-09-488-725A-5909	Sequence 5909, Ap
64	160.5	4.2	1664	28	US-10-258-898A-5909	Sequence 5909, Ap
65	160.5	4.2	1664	28	US-10-286-897-5909	Sequence 5909, Ap
66	160.5	4.2	1702	1	PCT-US01-08631-38864	Sequence 38864, A
67	160.5	4.2	1702	1	PCT-US01-08631-38864	Sequence 38864, A
68	160	4.2	5303	37	US-60-173-464-23675	Sequence 23675, A
69	159	4.2	890	1	PCT-US01-14827-12944	Sequence 12944, A

70 158.5 4.2 1585 35 US-10-990-328-9765 Sequence 9765, Ap
71 158.5 4.2 4866 37 US-60-191-637-29941 Sequence 29941, A
72 158.5 4.2 4866 37 US-60-191-681-24058 Sequence 24058, A
73 158.5 4.2 5303 20 US-09-614-150-30390 Sequence 30390, A
74 158.5 4.2 5303 20 US-09-614-150A-30390 Sequence 30390, A
75 158 4.2 2768 20 US-09-614-150-31983 Sequence 31983, A
76 158 4.2 2768 20 US-09-614-150A-31983 Sequence 31983, A
77 158 4.2 2768 37 US-60-173-464-24710 Sequence 24710, A
78 158 4.2 2768 37 US-60-191-637-31555 Sequence 31555, A
79 158 4.2 2768 37 US-60-191-681-25036 Sequence 25036, A
80 157 4.1 866 27 US-10-170-205E-7942 Sequence 7942, Ap
81 157 4.1 866 37 US-60-455-444-7969 Sequence 7969, Ap
82 157 4.1 866 37 US-60-465-241-7969 Sequence 7969, Ap
83 157 4.1 897 1 PCT-US03-14450-45 Sequence 45, Appl
84 157 4.1 1772 30 US-10-437-963-143280 Sequence 143280, A
85 157 4.1 2766 25 US-09-964-956-62 Sequence 62, Appl
86 157 4.1 2766 31 US-10-514-150-6 Sequence 6, Appl
87 156.5 4.1 1454 27 US-10-170-205E-20817 Sequence 20817, A
88 156.5 4.1 1503 20 US-09-649-996-14 Sequence 14, Appl
89 156 4.1 888 1 PCT-US01-08631-44262 Sequence 44262, A
90 156 4.1 888 1 PCT-US01-08631-54269 Sequence 54269, A
91 156 4.1 888 1 PCT-US04-07412-1795 Sequence 1795, Ap
92 156 4.1 888 29 US-10-389-559-1795 Sequence 1795, Ap
93 156 4.1 888 37 US-60-365-264-603 Sequence 603, App
94 155.5 4.1 1801 30 US-10-438-246-17202 Sequence 17202, A
95 155 4.1 440 27 US-10-179-131-6540 Sequence 6540, Ap
96 154 4.0 2992 1 PCT-US03-26780-2188 Sequence 2188, Ap
97 154 4.0 3065 1 PCT-US03-26780-2257 Sequence 2257, Ap
98 154 4.0 4873 28 US-10-236-177-240 Sequence 240, App
99 154 4.0 4919 37 US-60-568-845-241 Sequence 241, App
100 154 4.0 5314 37 US-60-568-845-240 Sequence 240, App

ALIGNMENTS

RESULT 1
US-10-650-482-2
; Sequence 2, Application US/10650482
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; TITLE OF INVENTION: PHOSPHATASE SUBUNIT
; FILE REFERENCE: 5986/11712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; CURRENT FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-482-2

Query Match 100.0%; Score 3804; DB 32; Length 713;
Best Local Similarity 100.0%; Pred. No. 2.5e-237;
Matches 713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60

Qy 61 SYWTKLLSQLAPLPGLLQKVLWSQLFGGMFPTRWLDFAGVYSALRKALGKREKPAAPTA 120
Db 61 SYWTKLLSQLAPLPGLLQKVLWSQLFGGMFPTRWLDFAGVYSALRKALGKREKPAAPTA 120

Qy 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180

Qy 181 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 240
Db 181 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 240

Qy 181 WGVLLPSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLEVSYQN 240
Db 181 WGVLLPSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLEVSYQN 240

Qy 241 SDGNSEVVGGTTLTPSSCLREDHCHPOPLSAELIPASWGQCPPLSTEGLEPIHHLRMKR 300
Db 241 SDGNSEVVGGTTLTPSSCLREDHCHPOPLSAELIPASWGQCPPLSTEGLEPIHHLRMKR 300

Qy 301 LEFLOQANKGODLPTPDQNGYHSLREHSLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301 LEFLOQANKGODLPTPDQNGYHSLREHSLRMDPKHCRDNPQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTVEPLALREESPSSECPSEIPMEKEPGEGRISVVVDYVYLEGDLPI SARPA 420
Db 361 TEEKIELLTVEPLALREESPSSECPSEIPMEKEPGEGRISVVVDYVYLEGDLPI SARPA 420

Qy 421 CSNKLIDYILGGASSDLETSSDPGEDWDEAEDDGFDSOSSLSDSLQDPEGLHLWNS 480
Db 421 CSNKLIDYILGGASSDLETSSDPGEDWDEAEDDGFDSOSSLSDSLQDPEGLHLWNS 480

Qy 481 FCSVDPNPNQFTATIQTAAARI VPEEPSDSEKOLSGKSDLENSQSGLPETPEHSSGEE 540
Db 481 FCSVDPNPNQFTATIQTAAARI VPEEPSDSEKOLSGKSDLENSQSGLPETPEHSSGEE 540

Qy 541 DMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFOTSGENKGCRCDSKTPSEISVAISE 600
Db 541 DMWESSADEAESLKLWNSFCNSDDPNPLNFKAPFOTSGENKGCRCDSKTPSEISVAISE 600

Qy 601 CHTLLSCVKQLLSQESSECPDSVQORDVLSGGRHTRVKKVTFLEEVTEYVYISGDRKRG 660
Db 601 CHTLLSCVKQLLSQESSECPDSVQORDVLSGGRHTRVKKVTFLEEVTEYVYISGDRKRG 660

Qy 661 PWEEFARDGCRFKRIQETEDAIGYCLTFEHRERMFNRLOGTCFKGLNVLKQC 713
Db 661 PWEEFARDGCRFKRIQETEDAIGYCLTFEHRERMFNRLOGTCFKGLNVLKQC 713

RESULT 2

US-10-170-205E-23916
; Sequence 23916, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23916
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-23916

Query Match 99.9%; Score 3799; DB 27; Length 713;
Best Local Similarity 99.9%; Pred. No. 6.4e-297;
Matches 712; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEPGTGSRKRLGPRAGFRFWPPPPRRSQAGSSKFPPTPLGPENSGNPTLLSSAQPETRV 60

Qy 61 SYWTKLLSQLAPLPGLLQKVLWSQLFGGMFPTRWLDFAGVYSALRKALGKREKPAAPTA 120
Db 61 SYWTKLLSQLAPLPGLLQKVLWSQLFGGMFPTRWLDFAGVYSALRKALGKREKPAAPTA 120

Qy 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSSDPSVTSPLDWLEEGIHWOYSPDPLKLELKAKGSALDPAQAFLLEQOL 180

Qy 181 WGVLLPSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLEVSYQN 240
Db 181 WGVLLPSSQLSRLYSNRELSSPSGPNLQRIIDNFVSVYLLNPSYLDGPPRLEVSYQN 240

```
Db 181 WGVLLPSSLSQRLYSNRELSSPSGPNLQRIIDNFSVSVLLNPSYLDCCPRLVSQN 240
Qy 241 SDGSEVVGFTLTPSSCLREDHCHPQLSAELIPASWQCPLSTEGPLPEIHILRMKR 300
Db 241 SDGSEVVGFTLTPSSCLREDHCHPQLSAELIPASWQCPLSTEGPLPEIHILRMKR 300
Qy 301 LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301 LEFLQASKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Qy 361 TEEKIELLTTEVPLALEEESPSECPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPSECPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Qy 421 CSNKLIDYILGGASDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPEGLHLWNS 480
Db 421 CSNKLIDYILGGASDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPEGLHLWNS 480
Qy 481 FCSVDPNPNQFTATIQTAAIIVPEPSPSEKOLSGKSDLENSSQSGSLPETPEHSGEE 540
Db 481 FCSVDPNPNQFTATIQTAAIIVPEPSPSEKOLSGKSDLENSSQSGSLPETPEHSGEE 540
Qy 541 DWESSADAEASLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSEI VAISE 600
Db 541 DWESSADAEASLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSEI VAISE 600
Qy 601 CHTLLSCKVQLLGSQSECPDSVQDVLSCGRHTHVRKKVTFLEEVTEYIISGDEDRKG 660
Db 601 CHTLLSCKVQLLGSQSECPDSVQDVLSCGRHTHVRKKVTFLEEVTEYIISGDEDRKG 660
Qy 661 PWEEFARDGCRFQKRIQETEDAIGYCLTTFEHRERMFNRLOGTCFKGLNVLKQC 713
Db 661 PWEEFARDGCRFQKRIQETEDAIGYCLTTFEHRERMFNRLOGTCFKGLNVLKQC 713

RESULT 3
US-60-452-680-17655
; Sequence 17655, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60-452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17655
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-17655

Query Match 99.9%; Score 3799; DB 37; Length 713;
Best Local Similarity 99.9%; Pred. No. 6.4e-297;
Matches 712; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEPTGGSRKGLGRAGRFPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 1 MEPTGGSRKGLGRAGRFPPPPRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Qy 61 SYWTKLSQLLAPLGLLQKVLINSQFGMFPTRWLDFAGVYSALRALKREKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLLQKVLINSQFGMFPTRWLDFAGVYSALRALKREKPAAPTA 120
Qy 121 QKSLSSQLDSDPSVTPDLWLBEGIHQVSPDPLKLELKAKGASLDPAQAFLLEQQL 180
Db 121 QKSLSSQLDSDPSVTPDLWLBEGIHQVSPDPLKLELKAKGASLDPAQAFLLEQQL 180
Qy 181 WGVLLPSSLSQRLYSNRELSSPSGPNLQRIIDNFSVSVLLNPSYLDCCPRLVSQN 240
Db 181 WGVLLPSSLSQRLYSNRELSSPSGPNLQRIIDNFSVSVLLNPSYLDCCPRLVSQN 240
```

```
Qy 241 SDGSEVVGFTLTPSSCLREDHCHPQLSAELIPASWQCPLSTEGPLPEIHILRMKR 300
Db 241 SDGSEVVGFTLTPSSCLREDHCHPQLSAELIPASWQCPLSTEGPLPEIHILRMKR 300
Qy 301 LEFLQANKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Db 301 LEFLQASKGQDLPTPDQNGYHSLBEHSLLRMDPKHCRDNPQFVPAAGDIPGNTQES 360
Qy 361 TEEKIELLTTEVPLALEEESPSECPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPSECPSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Qy 421 CSNKLIDYILGGASDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPEGLHLWNS 480
Db 421 CSNKLIDYILGGASDLETSSDPGEWDDEAEADGDFDSSLSLSDLEQDPEGLHLWNS 480
Qy 481 FCSVDPNPNQFTATIQTAAIIVPEPSPSEKOLSGKSDLENSSQSGSLPETPEHSGEE 540
Db 481 FCSVDPNPNQFTATIQTAAIIVPEPSPSEKOLSGKSDLENSSQSGSLPETPEHSGEE 540
Qy 541 DWESSADAEASLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSEI VAISE 600
Db 541 DWESSADAEASLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDSTPSEI VAISE 600
Qy 601 CHTLLSCKVQLLGSQSECPDSVQDVLSCGRHTHVRKKVTFLEEVTEYIISGDEDRKG 660
Db 601 CHTLLSCKVQLLGSQSECPDSVQDVLSCGRHTHVRKKVTFLEEVTEYIISGDEDRKG 660
Qy 661 PWEEFARDGCRFQKRIQETEDAIGYCLTTFEHRERMFNRLOGTCFKGLNVLKQC 713
Db 661 PWEEFARDGCRFQKRIQETEDAIGYCLTTFEHRERMFNRLOGTCFKGLNVLKQC 713
```

RESULT 4

```
US-09-629-469A-18965
; Sequence 18965, Application US/09629469A
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAOBU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/09/629,469A
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-629-469A-18965

```

Query Match          99.7%; Score 3794; DB 20; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.6e-296;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRWPPFPFRSQAGSSKFTPLGPNSEGNPTLLSSAQPETRV 60
Db 1 MEGTGGSRKRLGPRAGFRWPPFPFRSQAGSSKFTPLGPNSEGNPTLLSSAQPETRV 60

Qy 61 SYWTKLSQLLAPLGLQKVLWSQLFGGMPTRWLDAGVYSALRAKGRKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLQKVLWSQLFGGMPTRWLDAGVYSALRAKGRKPAAPTA 120

Qy 121 QKSLSSQLDSSDPVSTPLDWEELGHIWQYSPDPLKLELKAKGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSSDPVSTPLDWEELGHIWQYSPDPLKLELKAKGSALDPAQAFLLEQOL 180

Qy 181 WGVELLPSLQSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240
Db 181 WGVELLPSLQSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240

Qy 241 SDGNSEVVGFTLTPESSCLREDHCHPQLSABLIPASWQGCCPLSTEGLPPIHHLRMKR 300
Db 241 SDGNSEVVGFTLTPESSCLREDHCHPQLSABLIPASWQGCCPLSTEGLPPIHHLRMKR 300

Qy 301 LEFLQAKNGQDLPTPDQDNGYHSLREHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 301 LEFLQAKNGQDLPTPDQDNGYHSLREHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420

Qy 421 CSNKLDIYILGGASSDLTSSDPEGDWDEAEADDGFDSDSLSDSDLEQDPEGLHLWNS 480
Db 421 CSNKLDIYILGGASSDLTSSDPEGDWDEAEADDGFDSDSLSDSDLEQDPEGLHLWNS 480

Qy 481 FCSVDYPNPQNFATTIQTAAIRVPEPSDSEKDLGKSDLENSSQSGSLPTPEHSSGEE 540
Db 481 FCSVDYPNPQNFATTIQTAAIRVPEPSDSEKDLGKSDLENSSQSGSLPTPEHSSGEE 540

Qy 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTTPSEISVAISE 600
Db 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTTPSEISVAISE 600

Qy 601 CHTLLSCVKVLLGSESECPDSVQRDVLSGGRHTHVKKKVTFLFEVTEYYISGDEDRKG 660
Db 601 CHTLLSCVKVLLGSESECPDSVQRDVLSGGRHTHVKKKVTFLFEVTEYYISGDEDRKG 660

Qy 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTGCFKGLNVLKQC 713
Db 661 PWEEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTGCFKGLNVLKQC 713

```

RESULT 5

```

US-10-917-503-18965
; Sequence 18965, Application US/10917503
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: SAITO, KAORU
; APPLICANT: YAMAMOTO, JUNICHI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: WAKAMATSU, AI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: OTSUKI, TETSUJI
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
; FILE REFERENCE: 084335/0123
; CURRENT APPLICATION NUMBER: US/10/917,503

```

```

; CURRENT FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: US/09/629,469
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: JP 1999-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 1999-300253
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18965
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-917-503-18965

```

```

Query Match          99.7%; Score 3794; DB 35; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.6e-296;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRWPPFPFRSQAGSSKFTPLGPNSEGNPTLLSSAQPETRV 60
Db 1 MEGTGGSRKRLGPRAGFRWPPFPFRSQAGSSKFTPLGPNSEGNPTLLSSAQPETRV 60

Qy 61 SYWTKLSQLLAPLGLQKVLWSQLFGGMPTRWLDAGVYSALRAKGRKPAAPTA 120
Db 61 SYWTKLSQLLAPLGLQKVLWSQLFGGMPTRWLDAGVYSALRAKGRKPAAPTA 120

Qy 121 QKSLSSQLDSSDPVSTPLDWEELGHIWQYSPDPLKLELKAKGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSSDPVSTPLDWEELGHIWQYSPDPLKLELKAKGSALDPAQAFLLEQOL 180

Qy 181 WGVELLPSLQSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240
Db 181 WGVELLPSLQSRLYSNRELSSPGPLNIQRIDNFSVSVLLNPSYLDLCPRLLEVSYON 240

Qy 241 SDGNSEVVGFTLTPESSCLREDHCHPQLSABLIPASWQGCCPLSTEGLPPIHHLRMKR 300
Db 241 SDGNSEVVGFTLTPESSCLREDHCHPQLSABLIPASWQGCCPLSTEGLPPIHHLRMKR 300

Qy 301 LEFLQAKNGQDLPTPDQDNGYHSLREHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 301 LEFLQAKNGQDLPTPDQDNGYHSLREHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360

Qy 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 361 TEEKIELLTTEVPLALEEESPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420

Qy 421 CSNKLDIYILGGASSDLTSSDPEGDWDEAEADDGFDSDSLSDSDLEQDPEGLHLWNS 480
Db 421 CSNKLDIYILGGASSDLTSSDPEGDWDEAEADDGFDSDSLSDSDLEQDPEGLHLWNS 480

Qy 481 FCSVDYPNPQNFATTIQTAAIRVPEPSDSEKDLGKSDLENSSQSGSLPTPEHSSGEE 540
Db 481 FCSVDYPNPQNFATTIQTAAIRVPEPSDSEKDLGKSDLENSSQSGSLPTPEHSSGEE 540

Qy 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTTPSEISVAISE 600
Db 541 DWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTTPSEISVAISE 600

Qy 601 CHTLLSCVKVLLGSESECPDSVQRDVLSGGRHTHVKKKVTFLFEVTEYYISGDEDRKG 660
Db 601 CHTLLSCVKVLLGSESECPDSVQRDVLSGGRHTHVKKKVTFLFEVTEYYISGDEDRKG 660

```

Qy 661 PWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTCFKGLNVLKQC 713
Db 661 PWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTCFKGLNVLKQC 713

RESULT 6

US-09-757-028-2348
; Sequence 2348, Application US/09757028
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM001
; CURRENT APPLICATION NUMBER: US/09/757,028
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 2660
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2348
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-028-2348

Query Match 99.6%; Score 3789; DB 22; Length 720;
Best Local Similarity 99.7%; Pred. No. 4.2e-296;
Matches 711; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPEGTGSRKRLGPRAGFRFWPPFFRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 8 MPEGTGSRKRLGPRAGFRFWPPFFRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 67
Qy 61 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFFTRWLDFAGVYSALRALKGRKPAAPTA 120
Db 68 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFFTRWLDFAGVYSALRALKGRKPAAPTA 127
Qy 121 QKSLSSQLQDSSDPSVTPDLWLEEGHWOYSPDLEKLEKAGSALDPAQAFLLEQQ 180
Db 128 QKSLSSQLQDSSDPSVTPDLWLEEGHWOYSPDLEKLEKAGSALDPAQAFLLEQQ 187
Qy 181 WGVLELLPSSLOSRLYSNRELSSGSPGLNIQRIIDNFSVSVYLLNPSYLDLCPFRLEVSQN 240
Db 188 WGVLELLPSSLOSRLYSNRELSSGSPGLNIQRIIDNFSVSVYLLNPSYLDLCPFRLEVSQN 247
Qy 241 SDGNSEVVGFTLTTPSSCLREDHCHPQLSABLIPASWQGCPLSTEGLPETIHLRMKR 300
Db 248 SDGNSEVVGFTLTTPSSCLREDHCHPQLSABLIPASWQGCPLSTEGLPETIHLRMKR 307
Qy 301 LEFLQAKNGQDLPTPDQNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 308 LEFLQAKNGQDLPTPDQNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 367
Qy 361 TEKIELLTTEVPPLALEEBSPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 368 TEKIELLTTEVPPLALEEBSPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 427
Qy 421 CSNKLIDYILGGASSDLETSSDPEGDWDEAEDDGFSDSSLSDSLQDPEGLHLWNS 480
Db 428 CSNKLIDYILGGASSDLETSSDPEGDWDEAEDDGFSDSSLSDSLQDPEGLHLWNS 487
Qy 481 FCSVDPNPNQFTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 540
Db 488 FCSVDPNPNQFTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 547
Qy 541 DDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600
Db 548 DDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 607
Qy 601 CHTLLSCVKQLLGQSECPDSVQDVLSCGRHHTVKKKVTFLVEETYYISGDEDRKG 660
Db 608 CHTLLSCVKQLLGQSECPDSVQDVLSCGRHHTVKKKVTFLVEETYYISGDEDRKG 667

Qy 661 PWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTCFKGLNVLKQC 713
Db 668 PWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMFNRLOQTCFKGLNVLKQC 720

RESULT 7

US-10-222-911-2348
; Sequence 2348, Application US/10222911
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM001CIN
; CURRENT APPLICATION NUMBER: US/10/222,911
; CURRENT FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: 09/757,028
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 2660
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2348
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-222-911-2348

Query Match 99.6%; Score 3789; DB 28; Length 720;
Best Local Similarity 99.7%; Pred. No. 4.2e-296;
Matches 711; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPEGTGSRKRLGPRAGFRFWPPFFRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 60
Db 8 MPEGTGSRKRLGPRAGFRFWPPFFRRSQAGSKFPTPLGPENSGNPTLLSSAQPETRV 67
Qy 61 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFFTRWLDFAGVYSALRALKGRKPAAPTA 120
Db 68 SYWTKLSQLLAPLPGLLQKVLWSQLFGGMFFTRWLDFAGVYSALRALKGRKPAAPTA 127
Qy 121 QKSLSSQLQDSSDPSVTPDLWLEEGHWOYSPDLEKLEKAGSALDPAQAFLLEQQ 180
Db 128 QKSLSSQLQDSSDPSVTPDLWLEEGHWOYSPDLEKLEKAGSALDPAQAFLLEQQ 187
Qy 181 WGVLELLPSSLOSRLYSNRELSSGSPGLNIQRIIDNFSVSVYLLNPSYLDLCPFRLEVSQN 240
Db 188 WGVLELLPSSLOSRLYSNRELSSGSPGLNIQRIIDNFSVSVYLLNPSYLDLCPFRLEVSQN 247
Qy 241 SDGNSEVVGFTLTTPSSCLREDHCHPQLSABLIPASWQGCPLSTEGLPETIHLRMKR 300
Db 248 SDGNSEVVGFTLTTPSSCLREDHCHPQLSABLIPASWQGCPLSTEGLPETIHLRMKR 307
Qy 301 LEFLQAKNGQDLPTPDQNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 308 LEFLQAKNGQDLPTPDQNGYHSLBEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 367
Qy 361 TEKIELLTTEVPPLALEEBSPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 420
Db 368 TEKIELLTTEVPPLALEEBSPEGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA 427
Qy 421 CSNKLIDYILGGASSDLETSSDPEGDWDEAEDDGFSDSSLSDSLQDPEGLHLWNS 480
Db 428 CSNKLIDYILGGASSDLETSSDPEGDWDEAEDDGFSDSSLSDSLQDPEGLHLWNS 487
Qy 481 FCSVDPNPNQFTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 540
Db 488 FCSVDPNPNQFTATTQTAARIIVPEEPSDSKOLSGKSDLENSQSGSLPETPHSSGEE 547
Qy 541 DDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 600
Db 548 DDWESSADEAESLKLWNSFCNSDDPNPLNFKAPFQTSGENEKGCRDCKTSPESIVAISE 607

QY 601 CHTLLSKVQLGSGQSECPDSVQRDVLSGGRHHTVKKRVTFLEEVTEYISGDEDRKG 660
 Db 608 CHTLLSKVQLGSGQSECPDSVQRDVLSGGRHHTVKKRVTFLEEVTEYISGDEDRKG 667
 QY 661 PWEEFARDGCRFQKRIQETEDAIGCLTFEHRERMFNRLOGTCFGLNLVKQC 713
 Db 668 PWEEFARDGCRFQKRIQETEDAIGCLTFEHRERMFNRLOGTCFGLNLVKQC 720

RESULT 8
 PCT-US00-05883-941
 ; Sequence 941, Application PC/TUS0005883
 ; GENERAL INFORMATION:
 ; APPLICANT: Craig Rosen,
 ; APPLICANT: Steve Ruben,
 ; TITLE OF INVENTION: Human Colon Cancer Associated Gene Sequences and Polypeptides
 ; FILE REFERENCE: PA102PCT
 ; CURRENT APPLICATION NUMBER: PCT/US00/05883
 ; CURRENT FILING DATE: 2000-03-08
 ; EARLIER APPLICATION NUMBER: 60/124,270
 ; EARLIER FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 941
 ; LENGTH: 707
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (265)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (271)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (307)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 PCT-US00-05883-941

Query Match 95.3%; Score 3627; DB 1; Length 707;
 Best Local Similarity 98.8%; Pred. No. 5e-283;
 Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 23 PFFPRRSQAGSKFPTPLGPGNSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 82
 Db 17 PSVSRRSQAGSKFPTPLGPGNSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 76

QY 83 IWSQLFGCMFPTRWLDPAQVYSAALRAKGRKPAAPTAQKSLSSQLDSDSPSVTSPLDW 142
 Db 77 IWSQLFGCMFPTRWLDPAQVYSAALRAKGRKPAAPTAQKSLSSQLDSDSPSVTSPLDW 136

QY 143 LEEGIHWQYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVVELLPSSLSQRLYSNREIGS 202
 Db 137 LEEGIHWQYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVVELLPSSLSQRLYSNREIGS 196

QY 203 SPSPGLNQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTTPSSCLRE 262
 Db 197 SPSPGLNQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTTPSSCLRE 256

QY 263 DHCHPQPLSABLIPASWQGCPLSTEGPLPEIHHLRMKLEFLQQAASKQDXPTPDQNGY 322
 Db 257 DHCHPQPLSABLIPASWQGCPLSTEGPLPEIHHLRMKLEFLQQAASKQDXPTPDQNGY 316

QY 323 HSLBEHSLRLMDPKHCRDNTQFVPAAGDIPGNTQESTTEBKIELLTTEVPLALEESPS 382
 Db 317 HSLBEHSLRLMDPKHCRDNTQFVPAAGDIPGNTQESTTEBKIELLTTEVPLALEESPS 376

QY 383 EGCSPSEIPMEKEPGEGRISVVDVSYLREGDLPISARPACSNKLDYILGGASSDLETSSD 442
 Db 377 EGCSPSEIPMEKEPGEGRISVVDVSYLREGDLPISARPACSNKLDYILGGASSDLETSSD 436

QY 443 PEGEDWDEAEADDGDFSDSSLSDSLQDPEGLHLMNSFCSDVPYNQNFATTIQTAAARI 502
 Db 437 PEGEDWDEAEADDGDFSDSSLSDSLQDPEGLHLMNSFCSDVPYNQNFATTIQTAAARI 496

QY 503 VPPEPDSSEKDLGKSKDLENSQSGLPPTPEHSSGDEDDWESSADAESLKLWNSFCNS 562
 Db 497 VPPEPDSSEKDLGKSKDLENSQSGLPPTPEHSSGDEDDWESSADAESLKLWNSFCNS 556

QY 563 DDYNNPLNFKAPFOTSGENEGKGRDSTKTPSESVIAISECHTLLSKVQLGSGQSECPDS 622
 Db 557 DDYNNPLNFKAPFOTSGENEGKGRDSTKTPSESVIAISECHTLLSKVQLGSGQSECPDS 616

QY 623 VQRDVLSGGRHHTVKKRVTFLEEVTEYISGDEDRKGPWEEFARDGCRFQKRIQETEDA 682
 Db 617 VQRDVLSGGRHHTVKKRVTFLEEVTEYISGDEDRKGPWEEFARDGCRFQKRIQETEDA 676

QY 683 IGYCLTFEHRERMFNRLOGTCFGLNLVKQC 713
 Db 677 IGYCLTFEHRERMFNRLOGTCFGLNLVKQC 707

RESULT 9
 US-09-925-299-941
 ; Sequence 941, Application US/09925299
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA102
 ; CURRENT APPLICATION NUMBER: US/09/925,299
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 941
 ; LENGTH: 707
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (265)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (271)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (307)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-299-941

Query Match 95.3%; Score 3627; DB 24; Length 707;
 Best Local Similarity 98.8%; Pred. No. 5e-283;
 Matches 683; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 23 PFFPRRSQAGSKFPTPLGPGNSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 82
 Db 17 PSVSRRSQAGSKFPTPLGPGNSGNPTLLSSAQPETRVSYWTKLLSOLLAPLPGLLQKVL 76

QY 83 IWSQLFGCMFPTRWLDPAQVYSAALRAKGRKPAAPTAQKSLSSQLDSDSPSVTSPLDW 142
 Db 77 IWSQLFGCMFPTRWLDPAQVYSAALRAKGRKPAAPTAQKSLSSQLDSDSPSVTSPLDW 136

QY 143 LEEGIHWQYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVVELLPSSLSQRLYSNREIGS 202
 Db 137 LEEGIHWQYSPDDLKLELKAKGSALDPAQAFLLEQQLWGVVELLPSSLSQRLYSNREIGS 196

QY 203 SPSPGLNQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTTPSSCLRE 262
 Db 197 SPSPGLNQRIDNFSVSYLLNPSYLDCCFPRLEVSQNSDGNSEVVGFTLTTPSSCLRE 256

QY 263 DHCHPQPLSABLIPASWQGCPLSTEGPLPEIHHLRMKLEFLQQAASKQDXPTPDQNGY 322


```
Db 257 DHCHPOPLXAEIPXSWGCPPLSTEGLPBIHLRMKRLBFLQOASKGQDXTPTDQNGY 316
Qy 323 HSLREHSLRMDPKHCRDNPOTQVPAAGDI PGNTQESTBEKTELLTTEVPLALEBESPS 382
Db 317 HSLREHSLRMDPKHCRDNPOTQVPAAGDI PGNTQESTBEKTELLTTEVPLALEBESPS 376
Qy 383 EGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA CSNKLI DYILGASDLETSSD 442
Db 377 EGCPSSEIPMEKEPGEGRISVVDYSYLEGDLPI SARPA CSNKLI DYILGASDLETSSD 436
Qy 443 PEGEDWDEABDDGFDSDSLSDSLDRLQDPPEGLHLWNSFCVDPYNPONTATIQTAAARI 502
Db 437 PEGEDWDEABDDGFDSDSLSDSLDRLQDPPEGLHLWNSFCVDPYNPONTATIQTAAARI 496
Qy 503 VPEEPSDEKDLGSKSDLENSSGSLPETPEHSSGDEDDWESSADEASLKLWNSFCNS 562
Db 497 VPEEPSDEKDLGSKSDLENSSGSLPETPEHSSGDEDDWESSADEASLKLWNSFCNS 556
Qy 563 DDPYNPLNFKAPFOTSGENKGCGRDSKTPSESI VAISECHTLLSCKVQLLGSQSECPDS 622
Db 557 DDPYNPLNFKAPFOTSGENKGCGRDSKTPSESI VAISECHTLLSCKVQLLGSQSECPDS 616
Qy 623 VORDVLSGGRHTHVKRKVTFLBEVTEYIISGDEDRKGPWEFARDGCRFQKRIQETEDA 682
Db 617 VORDVLSGGRHTHVKRKVTFLBEVTEYIISGDEDRKGPWEFARDGCRFQKRIQETEDA 676
Qy 683 IGYCLTFEHRERMPNRLOQTCFGLNLVKQC 713
Db 677 IGYCLTFEHRERMPNRLOQTCFGLNLVKQC 707
```

```
RESULT 10
US-10-650-482-4
; Sequence 4, Application us/10650482
; GENERAL INFORMATION:
; APPLICANT: Ron, David
; TITLE OF INVENTION: METHODS OF SCREENING TEST COMPOUNDS USING GADD34L, AN eIF2alpha-S
; FILE REFERENCE: 5986/1L712-US1
; CURRENT APPLICATION NUMBER: US/10/650,482
; PRIOR FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2003-08-28
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-650-482-4
```

```
Query Match 58.4%; Score 2223; DB 32; Length 698;
Best Local Similarity 65.0%; Pred. No. 1.4e-169;
Matches 456; Conservative 53; Mismatches 174; Indels 18; Gaps 7;

Qy 1 MEPTGSGRRLGPRAGFRWPPPPRRSQAGSSKPTPLGPENSGNPTLLSSAQPTRV 60
Db 1 METGTHRRKRPGLRGSWFRLPFL-RRSHACSEFPFPPSRQPNGN-----SALPERRT 54

Qy 61 SYWTKLLSQLLAPLPGLLQKVLVWSQLFGGMFPTRWLDFAGVYSALRALKGRSKPAAPTA 120
Db 55 RYWTKLLSQLLAPLPSLFQKLLWSQLSGGLIPTRWLDFAASYSALRASRGRESDAPTV 114

Qy 121 QKSLSLSQLD-SSDPSTVSPLDWLEEGHWOYSPDLKLELKAKGSALDPAQAFLLEQQ 179
Db 115 QKSLSYTAAGLFAKTRVVTALARGGTPAVLVLRLEVKLKAQERALDSPAAPTFLEQQ 174

Qy 180 LWGVLELPSLSQSLYNSRELSSGSPGPIANTIDNFSVSVYLLNPSYLDGCFRLEVSQ 239
Db 175 LWGVLELPSLSQAGLVSHRELDSSSGPLSVQSGNFKVSVYLLNPSYLDYLPQUGURCQ 234
```

```
Qy 240 NSDGNSEVVGFTLTPDESSCLREDHCHPOPLXAEIPASWQGCCPPLSTEGLPBIHLRMK 299
Db 235 SSAGGQGVFGFTLTPDESSCLREDHCHPOPLXAEIPASWQGCCPPLSTEGLPBIHLRMK 294
Qy 300 RLEFLQOANKGQDLPPTDQNGYHSLREHSLRMDPKHCRDNPOTQVPAAGDI PGNTQ 359
Db 295 WLVEL-OPNQOQLPFLTDQNGYHSLREHSLRMDPKHCRDNPOTQVPAAGDI PGNTQ 349
Qy 360 STEEKTELLTTEVPLALEBESPSGCPSEI PMEKEPGEGRISVVDYSYLEGDLPI SARP 419
Db 350 PIEKKPELVIEV-----SQSPQSSLPFCLEPVEKECEEDHTNATDLSDRGESLPVSTRP 404
Qy 420 ACSNKLI DYILGASDLETSSDPPEGEDWDEABDDGFDSDSLSDSLDRLQDPPEGLHLW 479
Db 405 VCSNKLI DYILGASDLETSSDPPEGEDWDEABDDGFDSDSLSDSLDRLQDPPEGLHLW 464
Qy 480 SFCSDVPYNPONTATIQTAAARI VPEEPSDEKDLGSKSDLENSSGSLPETPEHSSG 539
Db 465 SFHSVDYYPKQNTATIQTAAARI PRDPSDGTSGWSGCGV-GSCQEGPLPETPDHSSG 523
Qy 540 EDDWESSADEASLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCGRDSKTPSESI VAIS 599
Db 524 EDDWESSADEASLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCGRDSKTPSESI VAIS 583
Qy 600 ECHTLLSCKVQLLGSQSECPDSVQRDVLVSGGRHTHVKRKVTFLBEVTEYIISGDEDRK 659
Db 584 GHHTLLSCKAQLLESQEDNCPGCGLGEALAGERYTHIKRKVKVTFLBEVTEYIISGDEDRK 643
Qy 660 GPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMPNRLO 700
Db 644 GPWEFARDGCRFQKRIQETEDAIGYCLTFEHRERMPNRLO 684
```

```
RESULT 11
PCT-US00-35017A-963
; Sequence 963, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR FILING DATE: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-35017A-963
```

```
Query Match 21.4%; Score 813; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 9.9e-57;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 DPNPLNFKAPFOTSGENKGCGRDSKTPSESI VAISECHTLLSCKVQLLGSQSECPDSV 623
Db 4 DPNPLNFKAPFOTSGENKGCGRDSKTPSESI VAISECHTLLSCKVQLLGSQSECPDSV 63

Qy 624 QRDVLVSGGRHTHVKRKVTFLBEVTEYIISGDEDRKGPWEFARDGCRFQKRIQETEDA 683
Db 64 QRDVLVSGGRHTHVKRKVTFLBEVTEYIISGDEDRKGPWEFARDGCRFQKRIQETEDA 123

Qy 684 GYCLTFEHRERMPNRLOQTCFGLNLVKQC 713
Db 124 GYCLTFEHRERMPNRLOQTCFGLNLVKQC 153
```

```
RESULT 12
US-10-296-115-963
; Sequence 963, Application US/10296115
```

```

; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 963
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-963

Query Match      21.4%; Score 813; DB 28; Length 153;
Best Local Similarity 100.0%; Pred. No. 9.9e-57;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      564 DPNPLNFKAPQTSGENEKGRDSTPSESIVAISECHTLLSCKVQLLGSQSECPDSV 623
Db      4 DPNPLNFKAPQTSGENEKGRDSTPSESIVAISECHTLLSCKVQLLGSQSECPDSV 63

Qy      624 QRDVLSGGRHTHVKKKVTFLFEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI 683
Db      64 QRDVLSGGRHTHVKKKVTFLFEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI 123

Qy      684 GYCLTFEHRMFMNRLQTCFKGLNVLKQC 713
Db      124 GYCLTFEHRMFMNRLQTCFKGLNVLKQC 153

RESULT 13
PCT-US01-14827-15743
; Sequence 15743, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 15743
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-14827-15743

Query Match      13.4%; Score 511; DB 1; Length 101;
Best Local Similarity 95.0%; Pred. No. 1.3e-32;
Matches 95; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      614 SQSECPDSVQDVLSSGRHTHVKKKVTFLFEEVTEYYISGDEDRKGPWFEFARDGCRFQ 673
Db      2 SQSECPDSVQDVLSSGRHTHVKKKVTFLFEEVTEYYISGDEDRKGPWFEFARDGCSVQ 61

Qy      674 KRIQETDAIGYCLTFEHRMFMNRLQTCFKGLNVLKQC 713
Db      62 KRIQETDAIGYCLTFEHRMFMNRLQTCCKCKGLNVLKQC 101

RESULT 14
US-60-177-571-4491
; Sequence 4491, Application US/60177571
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,

```

```

; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000201
; CURRENT APPLICATION NUMBER: US/60/177,571
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 5082
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 4491
; LENGTH: 73
; TYPE: PRT
; ORGANISM: HUMAN
US-60-177-571-4491

Query Match      10.6%; Score 402; DB 37; Length 73;
Best Local Similarity 100.0%; Pred. No. 4.8e-24;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      641 VTFLEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI GYCLTFEHRMFMNRLQ 700
Db      1 VTFLEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI GYCLTFEHRMFMNRLQ 60

Qy      701 GTCFKGLNVLKQC 713
Db      61 GTCFKGLNVLKQC 73

RESULT 15
US-60-162-247-4518
; Sequence 4518, Application US/60162247
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000127
; CURRENT APPLICATION NUMBER: US/60/162,247
; CURRENT FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 5442
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 4518
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Human
US-60-162-247-4518

Query Match      10.5%; Score 399; DB 37; Length 73;
Best Local Similarity 98.6%; Pred. No. 8.4e-24;
Matches 72; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      641 VTFLEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI GYCLTFEHRMFMNRLQ 700
Db      1 VTFLEEVTEYYISGDEDRKGPWFEFARDGCRFQKRIQETEDAI GYCLTFDHRMFMNRLQ 60

Qy      701 GTCFKGLNVLKQC 713
Db      61 GTCFKGLNVLKQC 73

Search completed: September 16, 2005, 10:53:27
Job time : 189.419 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 10:42:43 ; Search time 86.4089 Seconds
(without alignments)
1348.829 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEFPTGSGRKLGRAGFRP.....RMFNRLLQGTCKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 760079 seqs, 163465437 residues

Total number of hits satisfying chosen parameters: 760079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Pending Patents_AA New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260.5	6.8	674	7	US-11-040-219-2
2	195	5.1	605	7	US-11-040-219-5
3	167	4.4	1596	7	US-11-060-005-4
4	160.5	4.2	1562	6	US-10-450-763-57852
5	160.5	4.2	1702	6	US-10-450-763-38864
6	158.5	4.2	1585	6	US-10-990-328A-9765
7	158.5	4.2	5303	7	US-11-097-143-30390
8	158	4.2	2768	7	US-11-097-143-31983
9	157	4.1	2766	6	US-10-877-346-62
10	156.5	4.1	1503	1	PCT-US05-18850-797
11	156	4.1	757	8	US-60-685-372-466
12	156	4.1	888	6	US-10-450-763-44262
13	156	4.1	888	6	US-10-450-763-54269
14	150.5	4.0	1391	6	US-10-450-763-38863
15	149	3.9	1857	1	PCT-US03-10870-2173
16	149	3.9	1857	7	US-11-009-554-21
17	148.5	3.9	1192	7	US-11-097-143-5718
18	148	3.9	1208	6	US-10-450-763-54526
19	147.5	3.9	951	8	US-60-685-372-462
20	147	3.9	3394	1	PCT-US03-10870-529
21	145.5	3.8	899	1	PCT-US04-37982-1140
22	145	3.8	1606	1	PCT-US05-18850-981
23	144.5	3.8	1582	6	US-10-535-571-39
24	143.5	3.8	1413	7	US-11-097-143-12927
25	142.5	3.7	1572	6	US-10-990-328A-11742

26	142	3.7	2343	6	US-10-330-773A-904
27	141.5	3.7	1044	7	US-11-051-720-1563
28	141.5	3.7	1521	7	US-11-051-720-1564
29	141	3.7	1911	7	US-11-097-143-9906
30	140.5	3.7	1714	6	US-10-450-763-32497
31	139.5	3.7	635	7	US-10-450-763-52460
32	139	3.7	1346	7	US-11-060-005-2
33	138.5	3.6	1714	6	US-10-450-763-36776
34	138.5	3.6	1976	7	US-11-097-143-35049
35	138	3.6	759	7	US-11-097-143-18597
36	138	3.6	1142	7	US-11-044-051-73
37	138	3.6	1155	6	US-10-450-763-54525
38	137.5	3.6	825	8	US-60-710-726-1424
39	137.5	3.6	1416	7	US-11-097-143-23289
40	137.5	3.6	3067	7	US-11-097-143-34878
41	137	3.6	1203	1	PCT-US03-10870-2174
42	137	3.6	1715	7	US-11-097-143-1059
43	136.5	3.6	532	6	US-10-950-095-28
44	136.5	3.6	1543	7	US-11-097-143-34911
45	136	3.6	904	6	US-10-450-763-52462
46	136	3.6	1049	6	US-10-540-634-8
47	136	3.6	1235	6	US-10-540-634-6
48	135.5	3.6	1735	1	PCT-US05-17105-6441
49	135.5	3.6	1780	7	US-11-097-143-41241
50	134.5	3.5	314	6	US-10-514-020-4685
51	134.5	3.5	1253	7	US-11-185-924-18
52	134.5	3.5	1253	8	US-60-710-726-546
53	134.5	3.5	1259	6	US-10-940-774A-10366
54	134.5	3.5	2342	8	US-60-664-936-1036
55	134.5	3.5	2342	1	PCT-US05-10454-230
56	134.5	3.5	2468	8	US-60-664-936-1039
57	134.5	3.5	2468	8	US-60-664-936-1046
58	134.5	3.5	2468	8	US-60-664-936-1047
59	134.5	3.5	2519	6	US-10-450-763-46995
60	134.5	3.5	2522	6	US-10-940-774A-10237
61	134.5	3.5	2622	6	US-10-450-763-36777
62	134	3.5	1049	7	US-11-051-720-1565
63	134	3.5	1439	7	US-11-124-368A-291
64	134	3.5	1526	7	US-11-051-720-1566
65	134	3.5	3012	8	US-60-710-726-1384
66	134	3.5	5002	7	US-11-097-143-17961
67	133.5	3.5	748	6	US-10-330-773A-394
68	133.5	3.5	896	1	PCT-US04-17965-2159
69	133.5	3.5	896	1	PCT-US04-17965B-2159
70	133.5	3.5	896	1	PCT-US04-17965C-2159
71	133.5	3.5	896	6	US-10-863-905-2159
72	133.5	3.5	896	6	US-60-643-717-3535
73	133.5	3.5	1086	1	PCT-US05-10211-3
74	133.5	3.5	1192	7	US-11-090-836-3
75	133.5	3.5	1192	7	US-11-090-846-3
76	133.5	3.5	1192	7	US-11-090-847-3
77	133.5	3.5	1192	7	US-10-330-773A-392
78	133.5	3.5	1669	6	PCT-US05-18850-765
79	133.5	3.5	2244	1	US-10-940-774A-6452
80	133	3.5	513	6	US-11-185-924-16
81	133	3.5	513	6	US-60-710-726-548
82	133	3.5	513	6	US-10-940-774A-10364
83	133	3.5	534	6	US-10-489-448-1881
84	133	3.5	1439	6	US-10-450-763-40959
85	132.5	3.5	2905	8	US-60-664-936-1044
86	132.5	3.5	2817	1	PCT-US04-37982-299
87	132.5	3.5	2870	1	PCT-US04-37982-297
88	132.5	3.5	2870	1	PCT-US04-37982-293
89	132.5	3.5	2871	6	US-10-471-571A-2922
90	132	3.5	974	6	US-10-990-328A-12859
91	132	3.5	2753	6	US-10-940-774A-7659
92	132	3.5	2753	6	US-10-940-774A-7660
93	131.5	3.5	372	6	US-10-914-020-5746
94	131.5	3.5	509	7	US-11-097-143-33774
95	131.5	3.5	808	1	PCT-US03-10870-722
96	131.5	3.5	2858	7	US-11-097-143-40242
97	131.5	3.5	3060	7	US-11-097-143-984
98	131.5	3.5			Sequence 904, App
					Sequence 1563, Ap
					Sequence 1564, Ap
					Sequence 9306, Ap
					Sequence 32497, A
					Sequence 52460, A
					Sequence 2, Appli
					Sequence 36776, A
					Sequence 35049, A
					Sequence 18597, A
					Sequence 73, Appl
					Sequence 54525, A
					Sequence 1424, Ap
					Sequence 23289, A
					Sequence 34878, A
					Sequence 2174, Ap
					Sequence 1059, Ap
					Sequence 28, Appl
					Sequence 34911, A
					Sequence 52462, A
					Sequence 8, Appli
					Sequence 6, Appli
					Sequence 6441, Ap
					Sequence 41241, A
					Sequence 4685, Ap
					Sequence 18, Appl
					Sequence 1036, Ap
					Sequence 1038, Ap
					Sequence 230, App
					Sequence 1039, Ap
					Sequence 1046, Ap
					Sequence 1047, Ap
					Sequence 46995, A
					Sequence 10237, A
					Sequence 36777, A
					Sequence 1565, Ap
					Sequence 291, App
					Sequence 1566, Ap
					Sequence 1384, Ap
					Sequence 17961, A
					Sequence 394, App
					Sequence 2159, App
					Sequence 2159, Ap
					Sequence 2159, Ap
					Sequence 2159, Ap
					Sequence 3355, Ap
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 392, App
					Sequence 765, App
					Sequence 6492, Ap
					Sequence 16, Appl
					Sequence 548, App
					Sequence 10364, A
					Sequence 1081, Ap
					Sequence 40959, A
					Sequence 1044, Ap
					Sequence 299, App
					Sequence 237, App
					Sequence 293, App
					Sequence 232, Ap
					Sequence 12859, A
					Sequence 7659, Ap
					Sequence 7660, Ap
					Sequence 5746, Ap
					Sequence 33774, A
					Sequence 722, App
					Sequence 40242, A

Db 382 EDKEDSEALGEAES-----DHPGSH-PDRAHFRGWGYRPGKETEEBAEDWG 431
Qy 588 SKTPSESVISECHTLLSCVKQLGSG-----ESECPSDVSQDVLGGRHHT----- 635
Db 432 EAECPFRVAI-----YVGBKPPPPWAPPLRLQLRKLKRPETPHDPPDET 480
Qy 636 -VREKVTFLVEEVTYIS-----GDEDRKGPWEPFARDGCRFQKRIQETEDAIYCLTF 689
Db 481 PLKARKVRFBKVTVHFLAVWAGPAQAARQGPWELARDRSRPARRITQAQEELSCLTP 540
Qy 690 EHRERMPNRLQCTCFKGLNLVKQ 712
Db 541 AARARAWARLNPPLAPALATQ 563
RESULT 3
US-11-060-005-4
; Sequence 4, Application US/11060005
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A-A 070156 0597
; CURRENT APPLICATION NUMBER: US/11/060,005
; CURRENT FILING DATE: 2005-02-17
; PRIOR APPLICATION NUMBER: 09/902,432
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-060-005-4

Query Match 4.4%; Score 167; DB 7; Length 1596;
Best Local Similarity 19.8%; Pred. No. 0.58;
Matches 151; Conservative 112; Mismatches 286; Indels 214; Gaps 35;
Qy 6 GGSERKL-----GPR-----AGRFWPPPPFRSSQAGSSKEPT 38
Db 673 GSSKKRARKASSDDEGPRTLGGDSHRAEASKDKKAGTDVAPASTQEQDQAGSSPE 732
Qy 39 PLGPENSGNPTLLSSAQPETRVSYWTKLLSLLAPLPGLLQV-----LIWSQLFGGMF 92
Db 733 PAGSPSEGE-----GVSTW-SFVKELVTPRKKSKEKAEDSSVEQLSTEIE 780
Qy 93 PTR---WLDPAVGVYALRAKGRBKPAAPTAQKSL-----SLQDSSDPSVTS--PLDW 142
Db 781 PSREESWV-----SIKKFIPGRKRGADGKEQATVEDSGPVEINDDPNVPAVVELS- 833
Qy 143 LEEGIHWQSPDPLKELKAGSALDPAQAFLLEQLQWGLLPSSLSQLSLYSNRLGS 202
Db 834 -----EYNABE-REKMEAGNTELP-----QLLGA-----VYVSEELSK 866
Qy 203 SPSPGLNRIQIDNFSVVSYL--LNPSYLDCEP-----PRLEVSYQNSDGNSE 246
Db 867 TLVHTSVAVIDGTRAVTSVEERSPSMISASVTEPLEHTAGEAMPVEETEKDIIAET 926
Qy 247 VVGFTQITPSSCUREHCHPQ-PLSAELIPASWQGGCPPISTEGLEPEIHLHMRKRLFLQ 305
Db 927 PVLQTLL-PEGKADHDWVTSEVDFTSEAVTAT-ETSEALRTEEVTEASGAE-ETTDVMS 983
Qy 306 QANKQDLP-----TPDQ--DNVHSLREE-----HSLLRMDPKHCRDN----PTQVPA 349
Db 984 AVSQTLDSPDTEATPVQEVESGVLDTTEEBERTQAILQAVADKVKESQVPATQTVQR 1043

Qy 350 AGDIPGNQESTEEKIELLTTEVPLALAEESPEGCPSPSEIPMEKPFGEGRISVVDYSYL 409
Db 1044 TGSKALEKEVEEVEDSEVLASEKEKDVMPKGPVQEGAGAEHLAQSGSETQATPESLEVPE 1103
Qy 410 EGDLPISARAPACS-----NKLIDYILGASSDLETSSDPEG-----EDWDEAEODGFD 458
Db 1104 TADVDHVA--TCOVIKLQQLMEQAVAPSESETLTDSETNGSTPLASDSTADGTQOQDETID 1161
Qy 459 SDSLSLSDLEQDPEGLHLWNSFCSVDYPNPQNFATITOTAAIRIVPEEPS-----D 509
Db 1162 SQDSKATAVRQS-----QVTEEAATAQKEEFTLNNVPAQE 1200
Qy 510 SEKDLGKSLDNSQ---SGSLPETPEHSSGSED--DM---ESSADEAESLKLWNSFCN 561
Db 1201 EHGEPEGRDVLPTQOQELTAAAVPVLAKEVVGQEGEVDMLDGKVKEEQE----- 1250
Qy 562 SDDPYNPLNFKAPQTSGENEKGRDCKTPSEISVAISECHTLLSCVKQLGSGESECPCD 621
Db 1251 -----VFVHSGPNSQKAADVTDSE--VMGVAGCQEKESKEVOSLSLEEGEMET 1297
Qy 622 SVQRDVLGGRHTRVKKKVTFLVEEVTYYISGDEDRKGPWEE 664
Db 1298 DVE-----KEKRETKEQVSE---EGEQETAAPHE 1325
RESULT 4
US-10-450-763-57852
; Sequence 57852, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 57852
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1524)..(1556)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: eMATRIX, accession number PF00632C, p-value=3.302e-23, raw score
; OTHER INFORMATION: 20.66
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1562)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-value=4.3e-127, Pfam score of 435.6
US-10-450-763-57852
Query Match 4.2%; Score 160.5; DB 6; Length 1562;
Best Local Similarity 20.9%; Pred. No. 1.3;
Matches 161; Conservative 99; Mismatches 277; Indels 235; Gaps 40;
Qy 31 AGSSKEPTP-LGPNENSGNPTLLSSAQPETRVSYWTKLLSLLAPLPGL-LQKVLWSQL 87
Db 123 SGALRATTTSTVTWNSAAPIFKSGIADETVQGGSRRLISFSLDSFQAMGLKK----- 175
Qy 88 FGMWF--PTRWLDF--GVYSALRAL--KGREKPAAPTAQKSLSLQLDSSDPSVTSPL 140
Db 176 --GMFNPDPYLAISIQPGKHSIFPALPHHGER-----RSKIIG-----NTWNPI 219
Qy 141 DWLEEGIHWOYSPDP-LKLELKAGSALDPAQAQF-----LLEQQLMWGLLPSS 169

```

Db 220 -WQAEQSFVSLPTDVLIEVVDKFAKGRPIIKRFLGKLSMPVQRLRHAIGDRVVSYT 278
Qy 190 LOSRLYSNRELQ-----SSPSGPLNIQRIDNFSVVVLLNPSYLDLC 230
Db 279 LGRRLPTDHSVQLOFRFEITSSIHDPDEEISLSTEPESAQIQD--SPMNNLMESGSGE- 335
Qy 231 PPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLGSAELIPASWQGPCPLSTBGL 290
Db 336 -PRSEAPSESSWKPEQLGEGSVPRPGNQSIELSRPAEAAVITEAGDQGMVSGPEGA 394
Qy 291 PRIHHLRMKRLFLQOANKG-QDLPTPDQDNGYHSLBEHSLRLMDPKHCRDNPQTQFVPA 349
Db 395 GEL-----LAQVKDIQAPASABELAQDLGEEASALLLED----- 431
Qy 350 AGDIPGNTQESTBEKIELLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGRISVVDY 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEQEBEGDVST-----LEQEGRLQ---- 479
Qy 407 SYLEGDLPTISARPACSNKL--IDYILGASSDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVKRKRSP-CSLPVSELETVIASACGDPETPRTHYIRIHTLLHSMPSAQGSAE 536
Qy 453 EDGDFDSSLSDSLQDPEGLHLWNSFCSDVPYNPNQFTATQIQTAAIRIVPEPSDSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLSGKSDLENSQSGLPET--PEHSSG-----EEDWESSADAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPFSLANGAAQDQDTHPSTGESDS----- 616
Qy 560 CNSDDPNPLNFKAPFQTSGENEKGRDSTKTPSESAIVASCHTLLSKVOLLGQSESEC 619
Db 617 -----SPRQGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVQRDVLSGGR---HTHVKKRKVTFLFVETVYISGDEDRK----- 659
Db 661 P-----SCYNGNRFASFTRFSSVDSAKISESTVFSSQDDEEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-EEFARDGCRFQKRIQETEDAIGYCLTFEHRMFMNRLQGTG 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GFSNRREGEC 758

```

RESULT 5

```

US-10-450-763-38864
; Sequence 38864, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 38864
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1470)..(1498)
; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: eMATRIX, accession number PF00632B, p-value=3.700e-21, raw score
; OTHER INFORMATION: 18.45
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1257)..(1628)

```

```

; OTHER INFORMATION: HECT-domain (ubiquitin-transferase) domain identified by
; OTHER INFORMATION: Pfam, accession name HECT, E-value=5.5e-80, Pfam score of 279.1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1702)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-38864

```

Query Match 4.2%; Score 160.5; DB 6; Length 1702;

Best Local Similarity 20.9%; Pred. No. 1.4; Mismatches 277; Indels 235; Gaps 40; Matches 161; Conservative 99;

```

Qy 31 AGSKFPTP-LGPNNGNPTLLSSAQPETRVSVYWT-KLLSQLLAPLPGL-LQKVLWSQL 87
Db 123 SGALRAITPSTVTKNSAAPFKSIGADETVQGGSRRLISFSLSDFOAMGLK----- 175
Qy 88 FGMFNP--PTRWLQFA-----GVYSALRAL--KGRKPAAPATAQKSLSSQLQSDSDSVTSPL 140
Db 176 --GMFFNPDPYLKISIQGKHISFPALPHHQBGR-----RSKIIG-----NTVNP 219
Qy 141 DWLEEGIHQVSPDP-LKLELKAKGSAKDPAQAQF-----LLEQQLMGVLELLPSS 189
Db 220 -WQAEQSFVSLPTDVLIEVVDKFAKGRPIIKRFLGKLSMPVQRLRHAIGDRVVSYT 278
Qy 190 LOSRLYSNRELQ-----SSPSGPLNIQRIDNFSVVVLLNPSYLDLC 230
Db 279 LGRRLPTDHSVQLOFRFEITSSIHDPDEEISLSTEPESAQIQD--SPMNNLMESGSGE- 335
Qy 231 PPRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHPQPLGSAELIPASWQGPCPLSTBGL 290
Db 336 -PRSEAPSESSWKPEQLGEGSVPRPGNQSIELSRPAEAAVITEAGDQGMVSGPEGA 394
Qy 291 PRIHHLRMKRLFLQOANKG-QDLPTPDQDNGYHSLBEHSLRLMDPKHCRDNPQTQFVPA 349
Db 395 GEL-----LAQVKDIQAPASABELAQDLGEEASALLLED----- 431
Qy 350 AGDIPGNTQESTBEKIELLTTEVPLALEE---ESPSEGCPSSEIPMEKEPGRISVVDY 406
Db 432 -GEAPASTKEEPL-EEATTQSRAGREBEKEQEBEGDVST-----LEQEGRLQ---- 479
Qy 407 SYLEGDLPTISARPACSNKL--IDYILGASSDLET-----SSDPEGEDWDEEA 452
Db 480 --LRASVKRKRSP-CSLPVSELETVIASACGDPETPRTHYIRIHTLLHSMPSAQGSAE 536
Qy 453 EDGDFDSSLSDSLQDPEGLHLWNSFCSDVPYNPNQFTATQIQTAAIRIVPEPSDSEK 512
Db 537 EEDGAEESTLKDS---SEKQGL-----SEVDTV-----ADPSALEE 571
Qy 513 DLSGKSDLENSQSGLPET--PEHSSG-----EEDWESSADAESLKLWNSF 559
Db 572 D-----REEPEGATPGTAHPGSHGHPFSLANGAAQDQDTHPSTGESDS----- 616
Qy 560 CNSDDPNPLNFKAPFQTSGENEKGRDSTKTPSESAIVASCHTLLSKVOLLGQSESEC 619
Db 617 -----SPRQGDHSCGEC-DASCCSPSCYS-SSCYS-TSCYSSSCYSASCYS 660
Qy 620 PDSVQRDVLSGGR---HTHVKKRKVTFLFVETVYISGDEDRK----- 659
Db 661 P-----SCYNGNRFASFTRFSSVDSAKISESTVFSSQDDEEENSAPESVPSMQSPELD 715
Qy 660 -----GPW-EEFARDGCRFQKRIQETEDAIGYCLTFEHRMFMNRLQGTG 703
Db 716 PESTNGAGPWQDELAAPSGHVERSPGLESPVA-----GFSNRREGEC 758

```

RESULT 6

```

US-10-990-328A-9765
; Sequence 9765, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; OTHER INFORMATION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USSES THEREOF

```



```

; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31983
; LENGTH: 2768
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-31983

Query Match 4.2%; Score 158; DB 7; Length 2768;
Best Local Similarity 21.4%; Pred. No. 3.6;
Matches 139; Conservative 62; Mismatches 226; Indels 222; Gaps 28;

Qy 118 PTAQKSL-----SSLQDSSDPVTS-PLDWLEGIHWQVSPPLKLELKAGSAL 167
Db 1414 PTAEDLVPAFTPEIETSEFEVSTKPAVQGPPLPTLAPA-OPEKKPVDATST-1471

Qy 168 PPAQAFLLEQLQWGVLLPSSLQSLYSNRELSS-----PSGPLNIQRID 214
Db 1472 EPSAE-----VEKASGETSESNEIDAGASSTFPVVSADKTPSTKTEVEADD 1521

Qy 215 NFSVSYLLNFSYLCDFPRL-----EVSQNSDGNSE 246
Db 1522 KFTTVAPLAGDEESNLKLPQDIFEEAPVAVTTAAPS KODGEQKPVVEKEPIEDCQK 1581

Qy 247 VVGFTLLTPESCLREDHCHQPLSAELIPASQO-----GCPPLSTEGLPETHLBMK 299
Db 1582 PIEDTSTPTSS---ENBIEPESDRATTIAPSKKEPSPSTGAPT KDPAPFPSTDAPESD 1638

Qy 300 RLEFLQO-----ANKGQDLPT-----PDQ-----DNGY 322
Db 1639 ESKETPESEVPTTVAPAGEKIPTSSITPDREPTATSAVAKPDEDEVEKETETPTDAPA 1698

Qy 323 HSLEEHSLLRMDPKHC-----RDNPTQFVPAAGDIPGNT----- 357
Db 1699 SSEEDENSTDIQIPSEVPEKKPETAQ-TPEEGDIVGATAAPTTSDEVPVQVRLPEEVL 1757

Qy 358 ---QSTSEKI-----ELLTTEVPLA-----LEESPSREGC 385
Db 1758 EIPOPSTETGKQQDETAAFSIDRKEPVVTEIDEANTVAPISKDEKPEEKEPVQK 1817

Qy 386 PSSEIPMEKEPGEGRISVVDYSLYEGDLPISARPACSNKLDIYLGGASSDLETSSDPEG 445
Db 1818 PTGEPSSEEEKEKPIE-QDVS-TEGPVSTASEA-----GSTESSEEVKVPSTEG 1865

Qy 446 EDWDEAEDGFDSDSSLSDSLQEDQPEGLHLWNSFCVDPYFNQNFATTIQTAAIRIYPE 505
Db 1866 E-VAEKPED---KQPSSTAQAPVETIPE-----ISTELP-----AQDGDKPTSE 1905

Qy 506 EPSDEKDLGKSLDENSQSGLPTPE-----HSSGEEDDWESSADEASLKLWNSFC 560

```

```

Db 1906 APVDSDEDTAPSDEKIPSVSGEEVGEFVTTTSPQAAREDELKTPAESRPS----- 1957
Qy 561 NSDDPYNPLNFKAPFQTSGENEKGCRDSKTPSISVAISECHTLLSCKV 609
Db 1958 -STDKVPETEQP-----EDETKADETP-ESVTQVSDVATSTAPV 1997

RESULT 9
US-10-877-346-62
; Sequence 62, Application US/10877346
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/10/877,346
; CURRENT FILING DATE: 2004-06-25
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 2766
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-877-346-62

Query Match 4.1%; Score 157; DB 6; Length 2766;
Best Local Similarity 19.1%; Pred. No. 4.1;
Matches 171; Conservative 105; Mismatches 261; Indels 360; Gaps 47;

Qy 3 PGTG--GSRKRLGPRAGFRWPPFFPERSQAGSKPPTPLGPNSGNPTLLSSA-QPETR 59
Db 1164 PGTGWDGSSHL-----CSFGKSR---FVHPDSSTPTVABQVHQFES- 1203

Qy 60 VSYWTKLLSQLAPLPGLLQKVLWISQLFGCMFPFTRWLDFAGVYSALRALKGREKPAAPT 119
Db 1204 -----LSQVPSPRTSEPE-----SQGISKMKP-----PSQRCVSPREKASTPP 1241

Qy 120 AQKSLSLQLDSSDPVTSPLDWLEGIHWQVSPPLKLELKAGSALDPAA--QAFLE 177

```



```
Db 1242 -----DSSRA-----WAAPG-----DSSPSTRRIAVPMSTGAAPATAIPQASLVS 1281
Qy 178 QQ-----LWGVEL-LPSSIQ-----SRLYSNRELSSPGPLNIQR-----212
Db 1282 QERSGLSGPSKGLGTRELCLIPKSLKOGALLEDAPASGKMSHASSPSGVPATERTLSGS 1341
Qy 213 -----IDNFSVSYLLNPSYLCDFRLEPVSYQNSDGNSEVQFTLTPESSCLREDHC 265
Db 1342 PENPVTIDNFEBA-----SEARLSQSPQKADCRAGHGDTFESQPPGAGAGSSSH- 1391
Qy 266 HPOPLSAELIPA-----SWQGCCP-----LSTEGLEPEIH-----HIRMKLEFLQOANK-----309
Db 1392 HAQWVRSDQTSRPTGTGTGSPFPQWALQPSVLDSTHPDKHLAVNK-TFLNNYSRNFSN 1450
Qy 310 -----GQDLPTPDQNGVHSLSEEHSLLRMDPKHCRDNP-----TQFVP 348
Db 1451 PHEDSISLSGPGGSESESP-SSMYGNADSSS-----DPESLAEDFGAARNWNSPPLSP 1504
Qy 349 AAGDIPGNTQSTEEKIELLTTEVPLALAEESSEGCPSSEI-----PMEX 394
Db 1505 ESSPKGSS-ESDERIEICST-----DGCPTGPTVAPPPTQVALCPVLPVQ 1551
Qy 395 E-----PGEGRISVDYSYLEGDLPIGAR-----PACSN 423
Db 1552 RAVCKPVGDICERACFVPGASRTSIPDSSQPFSLDVSSEBPETWASINASQNHMPVCTE 1611
Qy 424 KLIDVI-----LGAGSSDLETS-----SDPE-----CED-WDEAEADDGF-----457
Db 1612 GIMDVTSSNMGDSQSQWTRHCRNAPPVLCNFMNDLGRDLDDEGAPKEGAAAASVM 1671
Qy 458 -----DSDSLSLD-----SDLEQDPEGLHL 477
Db 1672 RSVFALGAEKPKGEAVLADLHIAERGNLEDLQPKTISRRLTWFKEINKDSQSHL 1731
Qy 478 -----WNSFCSVDPNPQNFATIQTAARIVPEP-----SDSEKDLCKSDLEN---522
Db 1732 RSTSEKQSSMLALGPGSKAMVNTGHRKVTVPKSPRSQSKQENKDLPPKSPVETLGN 1791
Qy 523 -----SSQSGSLPETPE--HSSGEEDDWESSADEAE-SLKLWNS-----F 559
Db 1792 CQKPCSKPLKRLNSKGKASEVPVAISTKSRNDRHRTLPSPQASHKMFKAVSHRLHI 1851
Qy 560 CNSDDPNVPLNFKAPFOTSGENEGCR---DSKTPSEISVAISECHTLKCKVQLGSGE 616
Db 1852 ADQSEPKN-----TAGDTPKPPQCVPEKPPQAA-----LGLSLR 1885
Qy 617 SECPDSVQDVLS-----GGRHTVHKRKVTFLFEVTEYVISGDEDRKGP 661
Db 1886 TSASDTSIRTPTSPLTSPKLLPEQANSRFFM-----AVLESdTSCPTTSRPSRGP 1938
```

RESULT 10

```
PCT-US05-18850-797
; Sequence 797, Application PC/TUS0518850
; GENERAL INFORMATION:
; APPLICANT: CHIN, Lynda
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: DFN-059-2
; CURRENT APPLICATION NUMBER: PCT/US05/18850
; CURRENT FILING DATE: 2005-05-27
; NUMBER OF SEQ ID NOS: 1528
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 797
; LENGTH: 1503
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US05-18850-797
```

Query Match 4.1%; Score 156.5; DB 1; Length 1503;
Best Local Similarity 20.4%; Pred. No. 2.1;

```
Matches 147; Conservative 82; Mismatches 249; Indels 243; Gaps 33;
Qy 26 PRSQAGSKFPTLPGFENSGNPTLSSAQPETRVSYWTKLQLQALAPPELLQKVLWS 85
Db 826 PRRVP-----PDSLPTQGETQFTCLDIVPE-----DCLHQDISP-----DAVTPV 867
Qy 86 QLFQGMPTTWLDPAGVYSALRAKLGREKPAATAQKSLSSQLDSSDPSTVPLDWLEE 145
Db 868 EILSTDARTSHSLD-----NRQDSFG-----ESEETLRLTESDVLAD--DILAS 910
Qy 146 GIHWQSPDPLKLEKAKGSALDPAQAFLQOLMGVELLPSSLSQSRYSNRELSSPS 205
Db 911 RVSUGSLPLQGLQELNKPFSDEHSHR-RLEKNLEAVETLQO-----LNSKDAKEA 962
Qy 206 GPLNIQRIDNFSVSYLLNPSYLCDFRLEPVSYQNSDGNSEV-----247
Db 963 GLVSALSSDSTQSQS-LLEDLSAPFPASEPSLETDPDSLESVDVHEALDLSGHTPQKL 1021
Qy 248 -----VGFQTLTPESSCLREDHCHPQL-SAEILPAS-----WQGCCP-----LSTE 288
Db 1022 VPPDKPADSGYETENLESP---EWTLHPAPEGTADSEPATTDGCGHSGLPNPVIVISDA 1078
Qy 289 GLPEIHHLRMKLEFLQOANKGODLPTD-----ODNGVHSLSEEHSLLRMDPKH 338
Db 1079 G-----DHRGTEV-TPETFTAGSQSYRDSAYFS-----1107
Qy 339 CRDNPQTQVPAAGDIPGNTQSTEEKIELLTTEVPL---ALAEESPS--EGCPSSEIPME 393
Db 1108 --DNDSEPKRSEVPG-----TSPSALVLVQQLPEPVLPEQSPAAQDSCLEAR--K 1157
Qy 394 KPEGEGRISVDYSYLEGDLPIGARPACSNKLDIYLGGASSDLETSSDPEGEDW---448
Db 1158 SQPDESCLSLAHNS--SDLELATPEPAQT-----GVPPQVHPTEDEASSPWSVLA 1207
Qy 449 -----DEAEADD-----GFDSDSLSS-----DSLEQDPEGLHLWNSPCSDPN 488
Db 1208 ELSGDDFETQDRPCTLASTGNTNELLAYTNSALDKSLSSHEG-----1253
Qy 489 PONTATITQTAARIVPEEPSEDSKDLGSKDLSNSQSGSLPETPEHSSGEEDDWESSAD 548
Db 1254 -----PKLEPDIEGKY-LGKLGVSGMLDLSGDMDADEED-ENSDD 1293
Qy 549 EASLKLWNSFCNSDDPNPLNFKAPFOTSGENEGKCRDSTPSEISVAISECHTLK 608
Db 1294 SDEDLAFNLHLSSESEDETEHPVPIILSNEDGRHLR-----1331
Qy 609 VQLLGSESCPSDVQDVLSGGRHTVHKRKVTFLFEVTEYVISGDEDRKGPWEPEARD 668
Db 1332 -SULKPTAANAPDLPED-----WKEKKAIVFFDDVTYYLP-----DOETPTKELGPC 1379
Qy 669 G 669
Db 1380 G 1380
```

RESULT 11

```
US-60-685-372-466
; Sequence 466, Application US/60685372
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall D
; APPLICANT: Cousineau, Johanne
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: GeneMap of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 059908-5005-PR
; CURRENT APPLICATION NUMBER: US/60/685,372
```

```
; CURRENT FILING DATE: 2005-05-31
; NUMBER OF SEQ ID NOS: 2738
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 466
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-685-372-466

Query Match
Best Local Similarity 4.1%; Score 156; DB 8; Length 757;
Matches 118; Conservative 90; Mismatches 218; Indels 170; Gaps 30;

QY 3 PQTGSRKRLGRAGFRWPPFRSRQAGSKFPTPLGPENSGMPTLLSSAQBPTRVSY 62
DB 77 FGTLPST---PVTSF---PGIPDTLPFGSAPLEAPMTPTVDDSPQ----- 115
QY 63 WTKLLSQLAPLGLLQKVLINWSQLFGGMFPTP-----WLDPAFAG 101
DB 116 -KKMLGQKATPPPPPLSGELLKK---GSLPTSPRLVNESEMAVASGHLNSTGVLLVGG 171
QY 102 VYSALRALKGRKP---AAPTAKSL-----SIQLDSSDPSTVSPDLWLEEGIHW 149
DB 172 VLFMHGGEIQOTPTNTVAASPAASVSQATIVMMPALPAPSAAPAVST-----TESVA- 225
QY 150 QYSPDPLKLELKAKGSALDPAQAFLLEQOLMGVLLPSSLQSRLYSNRELGSPPSGPLN 209
DB 226 PVSQPDNCVPMBAVG---DPHTVTYSMDSS--EISMIINSIKKECFRS-GVAEAPVG--- 276
QY 210 IQRIDNFSVSYLLNPSVLDPCPRLE--VSYQNSDGNSEVVGFOFTLPSSCLREDHC-- 265
DB 277 -----SKAPSIDGKELDLAEKMDIAVSYTGEBLDPETVG-----DIITAIIDKVDD 323
QY 266 HPQPLSAELIPASWQGC-----PPLSTEGLPETHHLRMKRLFLQOANKGQDLPDPD--- 317
DB 324 HPEVLDAVAEAAALSCBENDDPQLPG-PWEHPIQ-----QERDKFVPLPAPEMTVK 375
QY 318 -----QNGYHSL---EEHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQSTQSTE--- 362
DB 376 QERLDFEETENKGIHELVDIREPSAEIKVEPAEPVLSGAEIVAGVVPATSMPEPLRS 435
QY 363 -----EKI-----ELLTTTEVPALRESP-----SEGCPSPSEIPEMEKPGEGRISVVDYS 407
DB 436 QDLDEELGSTAAGRIEADVAIGKDETPLTNVKTASPSML-----SPSHGSPNPIED-- 489
QY 408 YLEGDLPIASARPACSKLIDYILGASSDLETS-SDPEGDWDEAEADDGSDSSLSLD- 465
DB 490 -----PLEAETQHKFMSDSLSKEESGTIFGSIKDAPGED-----EEEDGVSEAAASLEEP 539
QY 466 -----SDLEQDPBGLHLWNSFCSVDYPNPQNFATTIQTAAARIVPEEPSPDSE 511
DB 540 KEEDQGEGLSEMNDNEPPVSE-----SDDGFSIHNATLQSHTLADSLPSSPASSQ 589

RESULT 12
US-10-450-763-44262
; Sequence 44262, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 44262
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-44262

Query Match
Best Local Similarity 4.1%; Score 156; DB 6; Length 888;
Matches 134; Conservative 64; Mismatches 209; Indels 176; Gaps 31;

QY 63 WTKLLSQLL-----APLPCLLQKVLINWSQLFGGMPTPRLWDPAGVYSALRALKGRKP 115
DB 40 WPLLMWQLLMLLVKBAQFLWVKDFLQUTNPLGP--PEPWSHSHSPF-----RESP 90
QY 116 AAPTAKSLSSQLDSSDPSTVSPDLWLEEGIHWQYSPDLKLE-----LKAAGSA----- 166
DB 91 HAPTL--PAUPWDFDHLGPSASSEMP-----APQOESTENLVFDFLDWDSAGELP 138
QY 167 LDPAQAFLLEQQLMGVLLPSSLQSRLL-YSNRELGSPPSGPLNIQRIIDNFSVSYLLNP 225
DB 139 LEP--EQLASQDQDLKDLSP--QERLPVSPKLLKDPARWLSLAEI--IGIIRQLSTP 191
QY 226 -----SYLD-----CPRLVSYQNSDGNSEVVG-----FOTLTPE----- 256
DB 192 QSQKQTLQNEYSTDTTPGSLPELKVKSDEPPGPPSEQVGPQSFHLEPETQNPETLEDI 251
QY 257 -SSCLREDHCHPOPLSAELIPASWQ---GCPPLSTEGLPETHHLRMKRLFLQOANKGQ 311
DB 252 QSSSLQQAEPALPQLLEEEPPSSMQEAPALPPSS-----NESLTLPNHEV 298
QY 312 DLPTPDQDNGYHSL-----EEHSLLRMDPKHCRDNPTQFVPAAG 351
DB 299 SVQPPGEDQAYYHLPNITVKPADVEVTTSEATNETESSQAQ-----QETPIQF-PEEV 351
QY 352 DIPGNTQSTEEKIELLTTTEVPALREB---SPSEGCPSSEI---PMEKE----- 395
DB 352 E-PEATQQAEPAPPPVPPMEHELISIEQQQPVQPS--TSREVSSTQQTQETPQCPPEHH 408
QY 396 -----PGEGRISVVDYSYL-----EGDLPISARPAC--SNKLIDY-----ILGASSDLE 438
DB 409 EVTVSPPGHHQTHLHDSFVSVKPPDVQLTIAAEPSEAEGVTSLVQQAETTLRSGSGNDIE 468
QY 439 TSSDPEGDWDEAEADDGSDSSLSLDLEQDPBGLHLWNSFCSVDYPNPQNFATTIQT 498
DB 469 PPAIQHG-----GPPLLSSESEAGFLAVQOETSFQSPSEFINNENPSPTQOE 515
QY 499 AARIVPEEPDSEKDLGKSDLENSSQSGSLP-----ETPEHS 536
DB 516 AA---AEHPQTABEGESSLTHQEAQAPQTPFPNPNVVAQPPHS 555

RESULT 13
US-10-450-763-54269
; Sequence 54269, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 54269
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-54269

Query Match
Best Local Similarity 4.1%; Score 156; DB 6; Length 888;
```

```
; OTHER INFORMATION: WW domain identified by Pfam, accession name WW, E-value=7.6e
; OTHER INFORMATION: -26, Pfam score of 99.3
US-10-450-763-38863

Query Match      4.0%; Score 150.5; DB 6; Length 1391;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 137; Conservative 87; Mismatches 213; Indels 255; Gaps 36;

Qy 63 WTKLLSOLL-----APLPGLLKVLWLSQLFCGMEPTRWLDPAQVYSALRALKGRKP 115
Db 40 WPLLMWQLLWLVKQAPLEWVKDPLQTSNPLGP--PEPWSSSHSHP-----RESP 90
Qy 116 AAPTAQKSLSLQDSSDPSTVSLDMLLEEGIHQYSPDLKLE-----LKAKGSA--- 166
Db 91 HAPTL--PADPWFDFHLGFSASSEMP-----APQOSTENLVFPFLDTWDSAGELP 138
Qy 167 LDPAQAFLLEQQLWGVLLPSSLSQRL--YSNRELGSPPSGPLNIQIDNFSVSYLLNP 225
Db 139 LEP--EQFLASQDLKDKLSP---QERLPVSPKLLKDPQARWSLAEI--IGIIRQLSTP 191
Qy 226 -----SYLD-----CPRLEVSQNSDGNSEVVG-----FOTLTP----- 256
Db 192 QSQKQTLQNEYSSTDTYPGSLPPELRVKSDEPPGPEQVGSQFPHLEPQNTLETEDI 251
Qy 257 -SSCLRDHCHPQLSAELIPASWQ-----GCPPLSTEGLEPEIHLRMKRLFFLQOANKGQ 311
Db 252 QSSSLQAEAPALQLLEEESSMQEAPALPPSS-----MESLTLPNHEV 298
Qy 312 DLPTPDQNGHSL-----EEHSLLRMDPKHCRDNPTQFVPAAG 351
Db 299 SVQPPGEDQAYHLFNITVKPADVEVTTSEATNETESSQAQ-----QETPIQF-PEEV 351
Qy 352 DIPGNTQESTEEKTELLTTEVPLALERE---SPSEGCPSSEI---PWEKE----- 395
Db 352 E-PSATQEAEPVPPVPMHELSIQQQQVQVQSE---TSREVESSTPQQETPGQPPEHH 408
Qy 396 -----PQEGRISVVVYSYL-----EGDLPISARPAC--SNKLIDY-----ILGASSDLE 438
Db 409 EVTVSPGHHQTHLSDPSVSVKPPDVQLTTAAPSAAEVGTSLVQQAETRLSGSGNDIE 468
Qy 439 TSSPPEGEDWDEAEADGFDSDSLSDLEQDPEGLHLWNSFCSDVPYNPQNTAIQT 498
Db 469 PPAIQHG-----GPPLLSSESEAGPLAVQOQTSFQSPPEPINNENPSPTQOE 515
Qy 499 AARIVPEPSKXDLGKSLDSENSQSGSLP-----ETREHS 536
Db 516 AA---AEHPQTAEGESLTHQEAAPQTPFPFNVVVAQPPEHS 555

RESULT 14
US-10-450-763-38863
; Sequence 38863, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CI21/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 38863
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (720)..  

; OTHER INFORMATION: WW/rsp5/WWP domain proteins domain identified by eMATRIX,
; accession number BL01159, p-value=8.568e-11, raw score of 13.85
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (705)..  

; FILE REFERENCE: 660088.465PC
```

```
; OTHER INFORMATION: WW domain identified by Pfam, accession name WW, E-value=7.6e
; OTHER INFORMATION: -26, Pfam score of 99.3
US-10-450-763-38863

Query Match      4.0%; Score 150.5; DB 6; Length 1391;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 137; Conservative 87; Mismatches 213; Indels 255; Gaps 36;

Qy 153 PPD-LKLELKAKGSALDPAQAF-----LLEQQLWGVLLPSSLSQSLRSLYNRELG 201
Db 8 PTDVLEIVKDKFAKSRPIIKRFLGKLSMPVQRLERHAIGDRVSVYTLGRRLLPTDH-- 64
Qy 202 SSPSGPLNIQIRINFVSVYLLNPSYLDPCPRLEVSQNSDGNSEVVG-----QT 252
Db 65 --VSGQLQFR---FRITS-SIHGYPN-----TTMLSSGRNDIPFPLPADDEEISLS 110
Qy 253 LTPSSSCLRE-----DCHCPOPLSAELIP----- 276
Db 111 TEPESAIOQSPNNLMESGSGPRSEAPRESSSWKPEQLGEGSVDPGPGNQIELSRPA 170
Qy 277 -----ASWQCPLPSTEGLEPEIHLRMKRLFFLQOANKG--QDLPTPDQNGYHSL 327
Db 171 EEAIVITEAGDQGMVSVGPEGAGEL-----LAQVQKDIQAPSAEELAEQLDLGE 220
Qy 328 EHSLLRMDPKHCRDNPTQFVPAAGDIPGNTQESTEEKIEILLTTEVPLALEE---ESPSEG 384
Db 221 EASALLIED-----GEAPASTKEBPLE--EBATTQSRAGREEEKEQBE 264
Qy 385 CPSEIPEMEKEPGEGRISVVVYSYLEGDLPIARPACSNKL--IDYILGASSDLET--- 439
Db 265 DVST-----LEQGEGRILQ-----LRASVKKRKRKRP-CSLPVSELETVIASAGDPETPRT 312
Qy 440 -----SSDPEGEDWDEAEADGFDSDSLSDLEQDPEGLHLWNSFCSDVPYNPQ 490
Db 313 HYIRIHLHSMPSAQGSAEEEDGAEESTLKDS---SEKQGL----- 354
Qy 491 NPTATTQTAARIVPEPSDEKOLSKSLDSENSQSGSLPET--PEHSSG----- 538
Db 355 ---SEVDTV---ADPSALEED-----RBEPEGATGTAHPGHSGHFPFLANGAA 399
Qy 539 -BEDDNESSADEAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENKGCSDKTPSEIVA 597
Db 400 QDGDTHPSTGESDS-----SPRQGDHSCGEC--DASCCSPSCYS 438
Qy 598 ISECHTLLSCKVQLLQSQSECPDSVQRDVLSGGR---HTHVKKRKVTFLEEVTEYISG 654
Db 439 -SSCYS--TSCYSSSCYSASCYSP-----SCYNGNRNFASHTRFSSVDSAKISESTVPSSQD 491
Qy 655 DEDRK-----CPW-BEFPARDGCRFKRIQETEDAI----- 683
Db 492 DEBEENSAPESVPDSMQSPELDPESTNGAGPWQDELAAPSGHVSERSPEGLESVPVAGPSNR 551
Qy 684 --GYCLTFEHRERMEVRLQTCFKGLNLVLAQC 713
Db 552 REGFTTADASRKQGFNM---TC-----LIRDC 575

RESULT 15
PCT-US03-10870-2173
; Sequence 2173, Application PC/TUS0310870
; GENERAL INFORMATION:
; APPLICANT: Mitokor, Inc.
; APPLICANT: Buck Institute
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465PC
```

```

; CURRENT APPLICATION NUMBER: PCT/US03/10870
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2173
; LENGTH: 1857
; TYPE: prt
; ORGANISM: Homo sapiens
PCT-US03-10870-2173

```

	Query Match	3.98;	Score 149;	DB 1; Length 1857;
	Best Local Similarity	20.76;	Pred. No. 7;	
	Matches 167; Conservative	80;	Mismatches 254;	Indels 304; Gaps 41;
Qy	21	WPPFFRRSAGSGSKFPTPLGP-----ENSGN-PTLLSSAQBPTRVSYWTKLISQLLAP	73	
Dd	294	WPCPALOARAGA---PAPGAPGEATPDQADNLVPVLGFPEKEORVOOLESELSKKRKK	350	
Qy	74	LPGLLKVKV-----LTW-----SQL	87	
Dd	351	CESLEQEARKKQRRCBELELRQAQWENARLVBENSRLSGRATEKQEVENAELRGOL	410	
Qy	88	FGGMFPTRLWDEFA-----GVYSALRALKGREKPAAPTAKSLSSQLDSDPSVTSPLDW	142	
Dd	411	LG---VTQERDSALRKQGLOSKLESUEVLKEMREVAQR-----	447	
Qy	143	LEEIGHWQYSPDOI.KLELKA.GSALDPAAQAF.LLEOQLGWGYELL.PSSLQSRLRYSNRELGS	202	
Dd	448	-RQGLEVEHQARLSIREKQEEVRRLQQAAQAEAREHEGA.VQLLESTLDSMQARVREEE	506	
Qy	203	SPSGPLNIQRIDNSVV-----SVLLNP.SVLDC.FPRLEVSV.QNSDGNSEVVGGTITPES	257	
Dd	507	QCR-----SQTEQFSLLAQBLQAFRLHPGPLD-----LLTSAL	539	
Qy	258	SCLREDHCHPQL-----SAELIPAS.WOGC.PPLSTE-----GLPETIHL	296	
Dd	540	DCGSLGDCPPPPCCCSIPQPCRGGPKDXLDLP.GSGPCRCTPKSEEPAPATLTGVPR---	595	
Qy	297	RMRKLEFLQANKGODLPTPDQNGYHSLEEHSLLRMDPKHCRDNTOVFPAAGDIPGN	356	
Dd	596	-----RTAKKAESL-----SNSHS-FSIHN-----SPKSC---PTPEVDTADEV---	631	
Qy	357	TQBSTEKIELLTT-----EYPLALEESPSEG--CPSSIPM-----	392	
Dd	632	-ELEADSVSLLPAAPEGSRGGARIQVFLARYSNPFEGPNENPEALPLTAGETIYYIG	690	
Qy	393	-----EKPEGGRISVVDYSYLE-----GDLPISARPACSKNLDIYI	429	
Dd	691	NMDBDGPFEGELMDGRRLGPSNFEVERVSDDDLTLTSLPPELLADLSHGSGPELS--FLSVG	748	
Qy	430	LGGASSDLETs---SDPEGEDWD.EEADDGFDSDSSLS.DSDLEDQPDRGLHLWNSFCSDVP	486	
Dd	749	GGGSSSGGGSGVRSQRPREE---EDAG--DELISLSPs-----PGEL---GEPPAVP	792	
Qy	487	YNPQNFTATTQTAARIY-PPEPSDSEKDLSG-----KSDLNSSQSGLSLPETPHSSGEE	540	
Dd	793	Y-PRLVLVQLAHSVVLAMEPPEQVELGHGFHCIVNGELRQALGPCAPPKA-----	843	
Qy	541	DDWESSADEABSLKLNWSFCNSDDPNPLNFKAPFQTS-GENEX-----GCEDSKTPS	592	
Dd	844	-----VLENLDLW-----AGPLHSVQALT.SRGSSDPLRCCLAVGARAGVVPs	886	
Qy	593	ESIVAISECHTLISCCKVOLL---GSQE-----SEC.PDSVQRDVLSGGRHTHVKKK	640	
Dd	887	QLRV-----HRLTATS.AEII.TWPGNS.NLAHAIYLNGECPPASPSTYWAIFCH-----	934	
Qy	641	VTFLEEVTEY--YISGEDRKGPMW	663	
Dd	935	--LRPGCTYQAQVEAQLPPQGPWE	956	

Search completed: September 16, 2005, 10:56:21
Job time : 88.4089 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 08:56:27 ; Search time 18.1914 Seconds
(without alignments)
3771.163 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MFPGTGGSRKRLGPRAGFR.....RMFNRLQGTCTGKGLNVLKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR 79.*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237	6.2	657	2 S10001	MyD116 protein - m
2	201	5.3	590	2 A56535	gadd34 protein - l
3	159	4.2	757	2 S68142	Probable transcrip
4	150.5	4.0	2282	2 T42717	DNA-binding protei
5	150	3.9	1062	2 T46444	hypothetical prote
6	149	3.9	647	2 S58225	skeletal muscle ab
7	148	3.9	1567	2 T03730	antigen containing
8	145	3.8	2774	2 A43359	microtubule-associ
9	144.5	3.8	3122	2 T17202	DNA-directed DNA p
10	143.5	3.8	574	2 S37762	225K protein - Bab
11	142.5	3.7	1621	2 A82255	hypothetical prote
12	142	3.7	406	2 S38170	SRP40 protein - ye
13	141.5	3.7	833	2 T22139	hypothetical prote
14	139	3.7	846	2 T25234	hypothetical prote
15	139	3.7	1346	2 A57376	probable regulator
16	139	3.7	1560	2 T30282	calcium-binding pr
17	139	3.7	5170	2 T15348	hypothetical prote
18	138.5	3.6	1271	2 A45555	glutamate rich pr
19	138	3.6	1311	2 G86471	unknown protein [i
20	137	3.6	1075	2 B96508	hypothetical prote
21	136.5	3.6	532	2 T06029	hypothetical prote
22	136.5	3.6	1807	2 S03124	vitellogenin A2 pr
23	136	3.6	772	2 A55004	transcription fact
24	136	3.6	1437	2 F96783	unknown protein F2
25	136	3.6	1736	2 T00391	hypothetical prote
26	135.5	3.6	534	2 T39903	serine-rich protei
27	135	3.5	1304	2 T19397	hypothetical prote
28	134.5	3.5	2218	2 B84683	hypothetical prote
29	134	3.5	1046	2 T29776	hypothetical prote

30	134	3.5	1830	2 A37981	microtubule-associ
31	133	3.5	854	2 T17288	hypothetical prote
32	133	3.5	917	2 B85057	hypothetical prote
33	133	3.5	933	2 S41539	fibrinogen-binding
34	133	3.5	1825	2 S13507	microtubule-associ
35	133	3.5	4307	2 T20721	hypothetical prote
36	132.5	3.5	966	2 S25365	CYC8 protein - yea
37	132.5	3.5	1447	2 S50918	DNA helicase TP81
38	131.5	3.5	1229	2 T46116	hypothetical prote
39	131.5	3.5	1435	2 A37793	erythrocyte-bindin
40	131	3.4	736	2 T00023	transcription fact
41	131	3.4	772	2 S62481	hypothetical prote
42	131	3.4	1048	2 S64758	SCD25 protein (ver
43	131	3.4	1772	2 A45532	major merozoite su
44	131	3.4	2761	2 T21064	hypothetical prote
45	130.5	3.4	1979	2 JW0059	microtubule-associ
46	130.5	3.4	2364	2 A56577	microtubule-associ
47	130.5	3.4	2416	2 T13825	adenomatous polypo
48	130	3.4	630	2 S29796	hypothetical prote
49	130	3.4	877	2 F90070	Clumping factor B
50	130	3.4	1262	2 T25168	hypothetical prote
51	130	3.4	1805	2 A34736	hypothetical prote
52	129.5	3.4	1108	2 A48508	cyclic-nucleotide
53	129	3.4	727	2 S18193	autoantigen NOR-90
54	129	3.4	727	2 JCS113	ribosomal transcri
55	129	3.4	764	2 S09318	transcription fact
56	129	3.4	764	2 JCS112	ribosomal transcri
57	129	3.4	1664	2 T18216	integrin-like prot
58	129	3.4	2342	2 T13412	hypothetical prote
59	128.5	3.4	194	2 A42749	hypothetical prote
60	128.5	3.4	1810	2 E88481	protein Cl6A3.2 [1
61	128.5	3.4	7160	2 T27935	hypothetical prote
62	128	3.4	756	2 T00367	hypothetical prote
63	128	3.4	989	2 D89852	fibrinogen-binding
64	128	3.4	990	2 I51618	nucleolar phosphop
65	128	3.4	1024	2 E86331	hypothetical prote
66	128	3.4	1087	2 S28282	hypothetical prote
67	128	3.4	1093	2 F88556	hypothetical prote
68	128	3.4	1684	2 JW0057	protein B0464.5a [
69	128	3.4	3924	2 S37431	gravin - human
70	127.5	3.4	490	2 G81723	ankyrin 2, neurona
71	127.5	3.4	568	2 I39411	conserved hypothet
72	127.5	3.4	820	2 T45685	AP-9 protein - hum
73	127.5	3.4	878	2 T17245	hypothetical prote
74	127.5	3.4	1463	2 T30290	hypothetical prote
75	127	3.3	589	2 S74668	AAS surface protei
76	127	3.3	781	2 T41551	hypothetical prote
77	127	3.3	1213	2 A58198	serine/proline-ric
78	127	3.3	1317	2 T03748	apoptosis associat
79	127	3.3	1472	2 A84470	hypothetical prote
80	127	3.3	1510	2 T33100	hypothetical prote
81	127	3.3	2327	2 T42630	aggreccan - bovine
82	126.5	3.3	889	2 T11742	egg sperm receptor
83	126	3.3	710	2 S30154	low-temperature-in
84	126	3.3	1300	2 T03166	probable immediate
85	125.5	3.3	1257	2 S28764	neurocan precursor
86	125.5	3.3	2132	1 A55182	aggreccan precursor
87	125	3.3	482	2 A44997	merozoite surface
88	125	3.3	636	2 T51893	related to Che-1 p
89	125	3.3	817	2 S53319	hypothetical prote
90	125	3.3	903	2 T47316	hypothetical prote
91	125	3.3	1804	2 T34518	nestin - golden ha
92	125	3.3	1812	2 I49350	breast/ovarian can
93	125	3.3	3375	2 T19821	hypothetical prote
94	125	3.3	5762	2 A41819	proline-rich pepti
95	124.5	3.3	968	2 T45746	hypothetical prote
96	124.5	3.3	1113	2 T47381	hypothetical prote
97	124.5	3.3	1153	2 F84468	hypothetical prote
98	124.5	3.3	1325	2 S62497	probable nucleopor
99	124.5	3.3	1355	2 S40022	spalt protein - fr
100.	124.5	3.3	1658	2 S55101	hypothetical prote

ALIGNMENTS

RESULT 1
 S10001
 Myd116 protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
 C:Accession: S10001
 R:Lord, K.A.; Hoffman-Liebermann, B.; Liebermann, D.A.
 Nucleic Acids Res. 18, 2823, 1990
 A:Title: Sequence of Myd116 cDNA: a novel myeloid differentiation primary response gene
 A:Reference number: S10001; MUID:90251472; PMID:2339071
 A:Accession: S10001
 A:Molecule type: mRNA
 A:Residues: 1-657 <LOR>
 A:Cross-references: UNIPROT:P17564; EMBL:X51829; NID:g53040; PIDN:CAA36128.1; PID:g53041

Query Match	6.2%	Score 237;	DB 2;	Length 657;
Best Local Similarity	21.8%;	Pred. No. 5.5e-07;		
Matches 169;	Conservative 81;	Mismatches 233;	Indels 292;	Gaps 35;
QY 53	SAQPTETVSVYWKLLS-QLLAPLPGLLQKVLVLSQLFGGFPPTRWLDFAGVYSALRAKLG 111			
DB 4	SPRQ-Q-VHLWRDAHNFYLLSPINGLLSRA--NSRLGPEVPEAWL----- 46			
QY 112	REKPAAPTAQKSLSSLDSDSPVTSPLDWLEGIHWQYSPPLKLELKAKGSALDPA 171			
DB 47	-----AKTVGDAQIE-----A 58			
QY 172	QAPLEQLQWGVLLPSSLSQRLYSNRELSSPSGPLNIQRIDNFSVSVYLLNPSYLD 231			
DB 59	AALITPTPVSG-NLTP-----HGTEEGSPESQSAQRL-----CL 94			
QY 232	PLREVSQNSDGNSEVVGFTLTPESSCLRDHCHQFPLSAELIPASWQGPCPLSTGLP 291			
DB 95	VEAESSPPETWGLSNVDYNA-KPGQDLREKEMERTAGKATLQAPAGLQAD----- 145			
QY 292	EIHILMRKRLEFLOQANKG---QDLTPDQDNGYHSLEEHSLR-----MDPKHCR 340			
DB 146	-----KRLGVVAREEGVAEPAYPTSQLEGGPAENEGDETVTKYQASAASTAFGYKP 198			
QY 341	DNPTQFVPAAGDIPGNTQESTEEKIELLITTEVPLALEEESPSEGCPS--EIPWKEPGE 398			
DB 199	STPVPLFGA-----EHQATEEK---GTE-NKADPSNPSGSHSRAWEYYSREPKQ 247			
QY 399	GRISVDVSYLEGDLPTISARPA-----CSNKLID-----YILGGAS 434			
DB 248	-----EGEAKVEAHRAQGQHPCRNAAEEGGPETTFVCTGNAFLKAWYRFGED 296			
QY 435	SDLTSSDPE-----GEDWDEAEEDDGFDSDSLSDSL 468			
DB 297	TEEDNSDSDSAEEDTAQTGATPHTSAFLKAWYRPGEDTEEE-----DSDSDSAEDT 350			
QY 469	EQDPEGHLWNPSVCSVDYPNPQNT-----ATIOTAARI---VDEEP 507			
DB 351	AQTGATPHT-SAFLLKAWYRPGEDTEENSDLSAEEDTAQTGATPHTSAFLKAWYRPG 409			
QY 508	SDSEKDLGKSDL-----ENSSQSGSLPET-----PEHSSGREDWES---- 545			
DB 410	EDTEEE---NSDLSAEEDTAQTGATPHTSPFLKAWYRPGEDTDDTEEESENVA 466			
QY 546	---SADRAESLKLWNSTFC-----NSDDPYNPLNFKAPPQTSGENEKCGRDSKTPSES 594			
DB 467	DSETADSSQSPCLQQRCLPCKEATKGRGEEP--PL-FQVAFYLPGEK-----PES 513			
QY 595	IVATSECHTLLSCVKQLLGS-----QSECEPDVSQVDVLSGRHHTHVKKVTFLEEVTEY 650			
DB 514	PWAPKPLRLQRLKRLFKATPRDQDPEIP-----LKARKVHFARKVTVH 558			
QY 651	YIS-----GDEDRKGPWEEFARDGCRFQKRIQETEDAIGYCLTFPEHRERMFNLQ 700			
DB 559	PLAWAGPAOAAARGPWEQFARDRSRFARIAQAEKLLGPLYLPDPSRARAWARLR 613			

RESULT 2
 A56535
 gadd34 protein - long-tailed hamster
 N:Alternate names: MYD116 protein homolog
 C:Species: Cricetus longicaudatus (long-tailed hamster)
 C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
 C:Accession: A56535
 F:Zhan, Q.; Lord, K.A.; Alamo Jr., I.; Hollander, M.C.; Carrier, F.; Ron, D.; Kohn, K.W.
 Mol. Cell. Biol. 14, 2361-2371, 1994
 A:Title: The gadd and MYD genes define a novel set of mammalian genes encoding acidic pro-
 A:Reference number: A56535; MUID:94187707; PMID:8139541
 A:Accession: A56535
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-590 >ZHA>
 A:Cross-references: UNIPROT:Q60465; GB:L28147; NID:g452489; PIDN:AAA36983.1; PID:g452490
 C:Genetics:
 A:Gene: Gadd34

	Query Match	5.3%;	Score 201;	DB 2;	Length 590;
	Best Local Similarity	23.9%;	Pred. No. 7.9e-05;		
	Matches 142;	Conservative 56;	Mismatches 210;	Indels 186;	Gaps 29;
Qy	245 SEVVGFQTTPSSCLREHCHPQPLSAEL-----IPASWQCPLSTEGLEPIHILRM	298			
Db	23 SPLMGF--LSRANSLRLVPEA-PEPFAETVIGADQIEADHPAPPL-----VPE-NHPPQ	74			
Qy	299 KRLEFLQANQGQDLPTPOQNGYHSLSEHSLLRMDPKHCRDNP-----TQFVP----	348			
Db	75 GRAESGTPKEAGKAGQCLIDVQANSPPETLGLSDDDKQGDGPREGQGRAHTAGLPILL	134			
Qy	349 -----AAGDIPGNTQESTEEKIELLTTEVPTALEEESPSEGCPSSE-----	389			
Db	135 SPGLQSADKSLGSEVAVAGEGVTE-----LAY-PTSHWEGCPSSEEDGETVKA	182			
Qy	390 -----IPMEKE-----PQEGRISVVDYSYLEGDLPISARPACSNKLIIDYILGGASSDL	437			
Db	183 FRASADSPGHKSTSVYCCEAEHQATEEKQETENKADPPSSPS-----GSHRAW	232			
Qy	438 ETSSDPEGEDWDE-----EASDDGPDSDSSLSDSDLQODPEGLHLWNSFCSV	484			
Db	233 EYCSKQGEADPFPRAGKYQLCONAEAESEAEKVSLSVSS-----GNAFLKA	282			
Qy	485 DPYNPQNFT-----ATIQTARIVPPEP-----SDSEKOLSGKSDLE	521			
Db	283 WTVRPGEDTDDDDSWGSAEBEGKALSSPTSEPHFLKAWVRPGEDTDD--DDSDWG	340			
Qy	522 NSSQSG---SLPPTPEHS-----SGE--EDDWESSADEAES-----L	553			
Db	341 SAESEKALSSPTSPEHDFLKAWVRPGEDTDDQDSDWGSAEKGLAQATFATPHTSAFL	400			
Qy	554 KLNSFCNSDDPN-----PLNFKA--PFQTSGENEKGRCDKTPSPSIVAISEC-HT	603			
Db	401 KTW-VCCPGGDTDDDCVVVPEDSBAADPKSPSHEAQGC-----LPFGQTEGLVAEHS	455			
Qy	604 LLSCKVQLGQSBS-----ECPDVSQVRDVL-----SGRRHTHVKKVKVTFLSEVTEY	650			
Db	456 LFQVAFYLPGEKAPAPWTAAPKPLRLRLRLTLRLTTQDDPETPLRARKVHSENVTVH	515			
Qy	651 YIS-----GDREDKGPWSEFARGCFQKRIQETEDAIGYCLTFEHRHMFNRL	699			
Db	516 FLVWAGPAQAARRGQWEOIARDSFARRIAQAEEKLGPYLTAPAFRAWARL	569			

```

RESULT 3
S68142
probable transcription factor SNAP - human
N:Alternate names: skeletal muscle abundant protein
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 16-Aug-2004
C:Accession: S68142

```

[illegible]

```

RESULT 5
T46444
hypothetical protein DKFZp434N1427.1 - human
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46444
R:Bloeker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wienan
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223032
A:Accession: T46444
A>Status: preliminary

```

[illegible]


```

Qy 313 LP-----TPDOONGYH-----SLEEBSLLRMDPKHCRDNPTQFVPAAGDIICNTQESTEEK 364
    |||
    |||
Db 1342 LPBKVAVVEQDLIIHQKDGALDENK-----PGRQODKTPQE 1378
    |||
    |||
Qy 365 IELLTTEVPLALBEEPSCEGSEIPEMKEGEGRISVVVDYVYLEGDLPLSARPACSNK 424
    |||
    |||
Db 1379 -----KGRDLDEKDTAAELDKGPEKELKDREDQQRAGP-----PAEKDK 1420
    |||
    |||
Qy 425 LIDYILGASDLETSDPGEWDWEAEADGDFDSGSSDSLEQDPQEGHLHWSFCSV 484
    |||
    |||
Db 1421 ASEQ-----RDTDLQOTQATEPRDRAERRDSE-EKOKSLERD----- 1458
    |||
    |||
Qy 485 DPYNPQNFTATIQTAAIV-----PE-----EPSDSEK--DLSGKSDLENSOGSLPET 532
    |||
    |||
Db 1459 -----RTPBEKORILVOERAPAEHSIPEPTQTDRAPRKGTDDKEQKEASEEKE- 1508
    |||
    |||
Qy 533 PEHSSGEDDW-----ESSADAESLKLWNSFCNSDDPYNPLNFKAPFOTSGENEGKCR 586
    |||
    |||
Db 1509 ---QVLEQKDMALGKEGTELDQARTAEKQDETILKE-----KTQGKSFVE 1553
    |||
    |||
Qy 587 DSKTPEESIV-----AISECHTLLSCKVOLIGSQSECPDSVQR-----D 626
    |||
    |||
Db 1554 DKTTTSKETVLDOKSAEKADSVEQDGAALAEKTRALGLEESPAEGSKAREQEKYKWEQD 1613
    |||
    |||
Qy 627 VLSGGRHTRVKRKVTFLEEVTEYVYISGDEDRKGPE 663
    |||
    |||
Db 1614 VVGWRETSPTREP-----VGGQKEPVPAAWE 1640
    |||
    |||

RESULT 9
T17202
DNA-directed DNA polymerase (EC 2.7.7.7) zeta chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17202
R:Van Sloun, P.P.H.; Romeijn, R.J.; Eken, J.C.J.
Mutat. Res. 433, 109-116, 1999
A:Title: Molecular cloning, expression and chromosomal localisation of the mouse
A:Reference number: Z18720; MUID:99202265; PMID:10102037
A:Accession: T17202
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3122 <V>A>
A:Cross-references: UNIPROT.Q61493; EMBL.AF083464; NID:G4079830; PID:G4079831;
A:Experimental source: strain 129/Ola; testis
C:Genetics:
A:Map position: 10
C:Keywords: nucleotidyltransferase

Query Match 3.88; Score 144.5; DB 2; Length 3122;
Best Local Similarity 19.69; Pred. No. 2.4;
Matches 142; Conservative 95; Mismatches 224; Indels 265; Gaps 37;

Qy 70 LLAPLPG-----LQKVLVLSQLFG-----GMPTTRMLDPAGYVSA 105
    |||
    |||
Db 13 MASPLPLGLDTCQSPLTQLPVKKVPV-VRVFGATPAGKTCCLHLHGIFPYLYVVDGY--- 68
    |||
    |||
Qy 106 LRALKGREKPAAPTAQKSL-----SLQDSDSDPSVTSP-----LDWLEGIHW 149
    |||
    |||
Db 69 -----GQPESYLSQWAFSIDBALNVALGNPSAQHVFKVSLVSGMPFYGYHEKERHF 122
    |||
    |||
Qy 150 Q-----YSPPLDK--LELKAGSGALD-----PAAQAFLEQQQLGWGVELL----- 186
    |||
    |||
Db 123 MKYLYNPAMVWKRICELQSGAIVMKCYQPHEAHPYLLQLFDYNDYGNWLNLAIVKP 182
    |||
    |||
Qy 187 -----PSSQLRYSNRELGLSPSG-----PLNQRIDNFSVSYLLNPSY 227
    |||
    |||
Db 183 RKARRKGNASHATCLFKHQLSGNSPAGTLFRWEDEIPSSLLLEGVEPLSTCELEVDVA 242
    |||
    |||
Qy 228 LDCFPRLVEYQNSDGNSEVVGQTLTPSESCULREDH-----CHPOPLSAELIPASWQ 281
    |||
    |||
Db 243 ADILNRLLDIEAQ-IGGNP---GLQAIWEDSKRRNRNNESSQISQPSQDCRFVPAT--- 295
    |||
    |||

```

```
Qy 282 CPPLSTEGLEPIHILRMKRL-EFLQANKGQDLP-TPDQNGYHSLREE---HSLLRMDP 336
Db 286 -----ESKQFQKRLQEVLKQNDFSVTLGSDVDYNGSQEFSAEILHSEI-LSP 344
Qy 337 KHCRRNPQFVPAAGDI---PQNTQESTEEKIELLTTEVPALAEBSPEGCPSEIPME 393
Db 345 EMLPCSPANMIEVHKDTOLSKGNTKHVEE--ALINEEAILNLIENTSQT----- 391
Qy 394 KEGGGRISVVDYSYLEGDLPIARPACSNKLIIDVILGASSD-----LE 438
Db 392 FOPLTORLUS-----ETVFMGSSPDLESIV-HLLAGLESDDGYGQGNRMPLPCHSFG 441
Qy 439 TSSDPGEGDWDEAEADGDFDSSLS-----DSDLQDDPEGLHLMNSFCSDVPYNPQN 491
Db 442 ESQNPQNSD-DEENEPQIEKEEMELSVVMQSRWSDIE----- 478
Qy 492 FTATIQTAARIVPREPSPSEKDLGSKDLENSSGSGSLPETPEHSSGEDWESSADAE 551
Db 479 -----EHCARKRSILCRNAHRSSSTEEDD-SSSEBEME 508
Qy 552 SLKLM--NSFC-----NSDDPNPLNFKAPFQTSGENEKGR-----DS 588
Db 509 ----WTDNSLLFANLISIPQLDGTADENSDNPLNNSRAHSSVATSKLSVRPSIFHKDA 564
Qy 589 KT--PSESIVAISECHTLLCKVKQLLG----SQSECPDSVQRDVLGGRTHVRKRKKVT 642
Db 565 ATLEPPSSAKITFECKHTSALSSHVLNKDLGTLDSQPNSTEK-----GRD-----NSVT 614
Qy 643 FLEEV 648
Db 615 FTKEST 620

RESULT 10
S37762
225K protein - Babesia bovis (fragment)
C:Species: Babesia bovis
C:Date: 19-May-1994 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: S37762
R:James, D.P.; Reduker, D.W.; Perryman, L.E.; McGuire, T.C.
Mol. Biochem. Parasitol. 52, 263-270, 1992
A:Title: A Babesia bovis 225-kilodalton protein located on the cytoplasmic side of the
A:Reference number: S37762; MUID:92319010; PMID:1377786
A:Accession: S37762
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-574 <JAS>
A:Cross-references: UNIPROT:Q00708; EMBL:N80466; NID:g155856; PIDN:AAA27791.1; PID:g155856

Query Match 3.8%; Score 143.5; DB 2; Length 574;
Best Local Similarity 20.3%; Pred. No. 0.27;
Matches 109; Conservative 78; Mismatches 194; Indels 157; Gaps 25;
Qy 57 ETRVSYWTKLSQLLAPLGLLQKVLWSQLFGGMFTRWLDFAGVYSALRAKGREKPA 116
Db 136 VKIRHYNRLVPLGLLQQLPDVLIKALKCNIVEGISPDVMLEI---EYLLKVNKNKPM 191
Qy 117 APTAKSLSSQLDSSD-----PSVTSPLDWLEEGHWHQVSPDLKLEKAKGSALDP 169
Db 192 SPEIRDTNDEMDTNDLSKPAQVTPPTPSP-----SITDE 226
Qy 170 AAQAFLEQQLGWVELLPSSQLSRLYSNRELGSPSGPLNRIQIDNFSVYLLNPSYLD 229
Db 227 VIQ---LPEYVDGVPVFFSHIDDEI-----IQLPD-----YEVSPIVLP 262
Qy 230 CFPRLVSYQNSDGNSEVVGFTLTTPESSCLURE-----DHCHPOPLSAELIPA 277
Db 263 IETIEGSESSSSSSDSEGEITLPEDSGVDPDDDDVIQLPEYVDSPIVLPETEI--- 319
Qy 278 SWQCPLPSTGLPEIHILRMKRLFLQANKGQDLPPTDQNGYHSLREEHSLLRMDPK 337
Db 320 --EGSESESSG-----SDSEGEITLIP-EDSGVE---EPDDDDVIQL-PE 357
```

```
Qy 338 HCRNPTQFVPAAGDIPGNTQESTEEKIELLTTE-VPLALEEBSPEGCPSSSE----- 389
Db 358 YV-DGPVVP-PSHID-----DEIIQLPDYEDSDPIVLPETIEIGSESSSSSSDSE 406
Qy 390 -----IPME---KEPGEGRISVVDYSYLEGDLPIARPACSNKLI---DY-----ILGAS 434
Db 407 GEEITLPEDSGVBEPPDDDDVIQLPEYV---DGPVVPFPHIDDEIIQLPDYEDSDPIVLP 462
Qy 435 SDLTSSDPGEGDWDEAEADGDFDSSLS--DSDLQDDPEGL-----HLMNSFCSV 484
Db 463 TEIGSESSSSSSGDSSEGEITLPEDSGVBEPPDDDDVIQLPEYVDPVVPVFFSHIDDEIIQL 522
Qy 485 DPYNPQNFTATIQTAARIVPREPSPSEKDLGSKDLENSSGSGSLPETPEHSSGEEDD 542
Db 523 PDYEDSDPIVLPETI-----ETEGSESSSSSSDSE-----GEEITLPEDSGVBEED 568

RESULT 11
A82255
hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82255
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1621 <HEI>
A:Cross-references: UNIPROT:Q9KTA5; GB:AE004181; GB:AE003852; NID:g9655454; PIDN:AAF9415
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0998
A:Map position: 1

Query Match 3.7%; Score 142.5; DB 2; Length 1621;
Best Local Similarity 20.5%; Pred. No. 1.3;
Matches 155; Conservative 100; Mismatches 278; Indels 223; Gaps 35;
Qy 33 SSKFPTP-----LGPENSGNPTL-----LSSAQ---PETRVSYWTKLSQLLAPLPG 77
Db 779 ANEFGTPQDEWDVDEDDSSFTLEGNAEELSAAEDDLPE-QTTATNETADELLADLAA- 836
Qy 78 LQKVLWSQLFGGMFTRWLDFAGVYSALRAKGREKPAAPTAQKSLSSQLDSSDPSVT 137
Db 837 -----QPQNTVDTSDDALAPDGLSQSVE 860
Qy 138 SPLD-----WLEEGHWHQVSPDPLKLEKAKGSALDPAA--- 171
Db 861 EPLTNDLLEPENDEPQLAEVTPSSAFDQEQVETEIEPESEPLAAEASNDSDLTALNE 920
Qy 172 ---QAFLEQQLGWVELLPSSQLSRLYSNRELGSPSGPLNRIQIDNFSVYLLNPSYL 228
Db 921 LDLPYTEEDVLADVQLEPAA-ESEVPDLELVNEPVEEAFTELDELDELDPYTEEDAL- 978
Qy 229 DCFPRLEVSYQNSDGNSEVVGFTLTTPESSCLREDHCHPOPLSAELIPASWQCQCPPLSTE 289
Db 979 -----ADAQLEPVAESEVEPELDLASE-----PABEEAFTELNLK 1013
Qy 289 GLPEIHH---LRMKRLEFLQANKGQDLP---PTPDQNGYHSLREEHSLLRMD- PKHCRD 341
Db 1014 DLPEYTEEDALAAQLESATESVESELELVSPAAREAFTELDE-----LDLPYTEE 1067
Qy 342 N---PTQFVPAAGDIPGNTQESTEEKIELL-----TTEVPLA--- 375
Db 1068 DALADSQLEPAA-----ESEVEPELELVSEPVTEEAFTELDELDELDPYTEEDALADAQ 1120
Qy 376 LEESPSGCPSPSEIPMEKPGEGGRISV-----VDY-SYLEGDLPI SAR--PACSNKLIDY 428
```

```
Db 1121 LEPAVESEVP--BLELASEPAEAEASTELNELDLPEYTEEDALADAQLEPAEAEV---1175
Qy 429 ILGGASDLTSSDPEGS-----DWDEEADDGFDSDSSLSDSLEQ-----DPE 473
Db 1176 -----ESELEASLEEBEPEPTNELDLPEYTEED-----ALADAQLEPAEAEVEPE 1224
Qy 474 GLHL-----WNSFCSVDYPNPQNFATIQTA-ARIVPEPSPDSEKDLGSKDLSNSQS 526
Db 1225 -LELATPAEBAEAPPELDELDPYTEEDALADAQLEPAEAEVEPELEASDLSEKETP 1283
Qy 527 GSLPE--TPEHSGE-----EDDESSADAEASLKLWNSFCNSDDPYNPLNPKA 573
Db 1284 TELDELDPYTEEDALADAQLEPAEAEVEPELEASVLEEDSFTELDELDPYTEYTEK 1343
Qy 574 PFQTSNGENKGRDKTP-----SESIVATSECHTLLSKVQLGSGEPCPSVQVORDVL 628
Db 1344 DALADAQLEPAEAEVEPELGDTELTIAQETESDALVADE--DLLASVES-AVDEVQPELL 1401
Qy 629 SGRHTRHKR--KVTFLVEEYTYISGDEDRKGPW 662
Db 1402 GATQDVPTQSLANKAFDEALHDLWLSNDPDKGKPF 1437

RESULT 12
SRP40 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR092c; protein YKR412a
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
A;Accession: S38170; S40645; S37702
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38170
A;Molecule type: DNA
A;Residues: 1-406 <BAL>
A;Cross-references: UNIPROT:P32583; EMBL:Z28317; NID:G486580; PID:G486581; MIPS:YKR092c
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Ji
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40645
A;Molecule type: DNA
A;Residues: 1-406 <BOU>
A;Cross-references: EMBL:X73541; NID:G450550; PIDN:CAA51946.1; PID:G450552
R;Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A;Description: Interactions between three common subunits of yeast RNA polymerases I and
A;Reference number: S37702
A;Accession: S37702
A;Molecule type: DNA
A;Residues: 1-399, 'N', 401-406 <LAL>
A;Cross-references: EMBL:L11275; NID:G295670; PID:G295671
C;Genetics:
A;Gene: SGD:SRP40
A;Cross-references: SGD:S0001800; MIPS:YKR092c
A;Map position: 11R

Query Match 3.7%; Score 142; DB 2; Length 406;
Best Local Similarity 23.4%; Pred. No. 0.21;
Matches 71; Conservative 27; Mismatches 121; Indels 84; Gaps 10;

Qy 370 TEVPLALEEPPSPGCPSSIEPMEKPGEGRISVVVYSYLEGLDLPISARPACSNKLIIDYI 429
Db 146 TEPSSSSSSSSSSSGSSSS---ESSEGS-----ESDSSSSSSSSSSSSSDSSDS 191
Qy 430 LGGASSDLETSSDPEGDWDEAEADDGFDSDSSLSDSLEQDPEGLHLWNSFCPSVDYPNP 489
Db 192 ESDSQSSSSSSSSSSSDSSSDSSSDSSSDSSSDSSSDSSSDSSSDSSSDSSSDSS 236
```

```
Qy 490 QNFATIQTAARIVPEPSPDSEKDLGSKDLSNSQSGLPTEPHSSGDEDDWESSADE 549
Db 237 -----DSDSSSDSSGSDSSSSSDSS-----DESTSSDSSDSSDSSDS 277
Qy 550 AESLKLWNSFCNSDDPYNPLNFKAPQTSNGENKGRDKSKTPSESIVATSECHTLLSKV 609
Db 278 GSSSELETKEATADES-----KABETPASSNE-----STPSAS----- 310
Qy 610 QLLGSGEPCPSVQVORDVLSGGRHTRH---VKRKVTF--LEEVTYYISGDEDRKGPWEEF 665
Db 311 --SSSANKLINIPAGTDEIKEGORKHFSRVDRSKINFEAWELTDNTYKG---AAGTWGEK 365
Qy 666 ARD 668
Db 366 ANE 368

RESULT 13
T22139
hypothetical protein F43G6.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
A;Accession: T22139; T22174
R;Swinburne, J.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19522
A;Accession: T22139
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-833 <WIL>
A;Cross-references: UNIPROT:Q20374; EMBL:Z50070; NID:e1519046; PIDN:CAA90402.1; GSPDB:GN
A;Experimental source: clone F43G6
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19526
A;Accession: T22174
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-833 <W12>
A;Cross-references: EMBL:Z83108; PIDN:CAB05512.1; GSPDB:GN00020; CESP:F43G6.9
A;Experimental source: clone F44E5
C;Genetics:
A;Gene: CESP:F43G6.9
A;Map position: 2
A;Introns: 21/1; 68/3; 122/3; 186/3; 381/3; 474/3; 641/2; 752/3

Query Match 3.7%; Score 141.5; DB 2; Length 833;
Best Local Similarity 18.1%; Pred. No. 0.6;
Matches 139; Conservative 106; Mismatches 251; Indels 273; Gaps 36;

Qy 6 GGSRKRLGPR-----AGPRFWPPFPFRRSQ-----GSSKPTPLGPNESGN--- 47
Db 50 GGGLDNIGENAELENTATQAKLRFPDPVWQKSSSDHVPASAEIPIPP--PNFGNGDA 107
Qy 48 -----PTLLSQAQETRVSVYWKLLSQLLAPLPGLLQKVLWSQLFG-----GMPTRWLDF 99
Db 108 SDSFKSFEAESPFLLKKSIGN-----CTDGAYNIWNTNFGISSLVPAAPTLLDLDF 157
Qy 100 AGV-----YSALRALKGRKPAAPTAQKSLSS---LQLDSSDPSVTSPLDWLEGI---- 147
Db 158 GALLPTFTTQATKEVKSQIPSPMPSSALTLEDCERMQMGKPNKPSLDVDAFKQLQLGSTFPVQ 217
Qy 148 --HWQYSPDPLKLELAKGKSGALDPAQAFLLEQLMGVELLPSSLQSRLYSNRELSSPS 205
Db 218 PQSAQSKHPLNARIAPGTPATASSQA-----LPTLPT 251
Qy 206 GPLNIQRIDNFSVSYLLNPSVLDLCPRLLEV--SYQNSDNGSEVGFQTLTPSSCLRED 263
Db 252 AALSLELE-----LQIMKEAQILKGRQV-----PSDWRED 283
Qy 264 H--CHQP-----LSAELIPASWQCPPLSTGLPEI-----HHLRMKRL----- 301
Db 284 NKFSHPPGFNQNVQPRMDPSLSPGMGHGMPSMGTSMPHGMPPQNMQLPRLPLNP 343
```

```

Qy 302 EFLQ-----QANQGDLPTP---DQNGYHSLLEEHSLLRMD--PKHCRRNP 343
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 344 QELPLIPVWFNAIINNIQLPMGVPPPPPFLFQLLNHYRNLPQLVHAMIQSSIPPNTFRQG 403
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 344 TGFV-PAAGDIGPNTQEETEEKIELLTTEVPALALEEESPSEGCPSSSEIPMEKEPGGRIS 402
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 404 PQSHPSGPHSPGNVRQKHSGMPSTRIYDLAID-----S 439
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 403 VVDY-SYLEGDLPISARPACSNKLIIDIYLGGASDLETSDPRGED----WDEEAEDDG 456
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 440 FAGVSYKERWLIR-----IQFI-----QCKGSGDPQVDDVVYVTWRDKQIANG 484
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 457 FUSDSSLSDSLDLEQDPGLHLWNSPFCSDVPNPNQFTATTQTAAIRIVPEPSPDSEKDLSG 516
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 485 WTAETKLEAEATEK-----KESSESQNDY---509
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 517 KSDLNSSOSGSLPTPBHSGEED-----DWESSADEAESLKNW-----S 558
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 510 ---LERISM-NYREMQERARERDKERQRERQERDRGEDKKLRQLTSLDKFATSLGLFS 565
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 559 FCNSDDPYNPLNFKAFFOTSGENEGKGRDSKTSPSEIVASECHTLILSCVKVLGGS----614
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 566 KSSTHNPRHVLMQAVESVDNQTKLSDES---RKIAVAKRLRTML---LRLEGALNL 619
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 615 -----QSECEPDVSQ-RDVLSGGRHTHVKKKTVLFEEVTEYYISGDE 656
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 620 MEVDLRRSSLPKSQFKDLSDEKQDEVEKRVTVIINE-----LMGDD 663
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
T25234
hypothetical protein T24D1.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25234
R;Cummings, P.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20001
A;Accession: T25234
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-846 <WIL>
A;Cross-references: UNIPROT:O02331; EMBL:Z81131; PIDN:CAB03423.1; GSPDB:GN00011
A;Experimental source: clone T24D1
C;Genetics:
A;Gene: CESP:T24D1.2
A;Map position: 1
A;Intons: 76/2; 125/2; 208/2; 258/2; 338/2; 369/3; 399/2; 517/1; 539/2; 749/2;
```

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 01:58:07 ; Search time 71.7548 Seconds
(without alignments)
5088.336 Million cell updates/sec

Title: US-10-650-482-2

Perfect score: 3804

Sequence: 1 MEPGTGSRKRLGPRAGFRF.....RMFNRLQGTGFKGLNLVKQC 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3795	99.8	713	2 Q6P156	Q6P156 homo sapien
2	3794	99.7	713	2 Q96SN1	Q96SN1 homo sapien
3	2419.5	63.6	697	2 Q8BFW3	Q8BFW3 m mus muscu
4	2177	57.2	407	2 Q658M2	Q658M2 homo sapien
5	1285	33.8	408	2 Q8PEGO	Q8PEGO mus musculus
6	1241	32.6	376	2 Q8C390	Q8C390 mus musculus
7	267	7.0	158	2 Q5VZB6	Q5VZB6 canarypox v
8	285.5	7.0	674	2 Q61A96	Q61A96 homo sapien
9	260.5	6.8	674	2 Q75807	Q75807 homo sapien
10	259.5	6.8	674	2 Q9NVU6	Q9NVU6 homo sapien
11	237	6.2	657	1 M116_MOUSE	P17564 mus musculus
12	217	5.7	578	2 Q7TQC2	Q7TQC2 rattus norv
13	214	5.6	578	2 Q6IN02	Q6IN02 rattus norv
14	201	5.3	590	2 Q60465	Q60465 cricetus
15	174	4.6	1444	2 Q9VTN2	Q9VTN2 drosophila
16	174	4.6	1514	2 Q8YSV5	Q8YSV5 drosophila
17	170	4.5	1095	2 Q8HCG6	Q8HCG6 homo sapien
18	169.5	4.5	353	2 Q6BQ12	Q6BQ12 debaryomyce
19	169	4.4	1596	2 Q9Z1F7	Q9Z1F7 rattus norv
20	169	4.4	10820	2 Q7SDK2	Q7SDK2 neurospora
21	167	4.4	1596	2 Q62766	Q62766 rattus norv
22	162	4.3	4915	2 Q6CJ86	Q6CJ86 kluyveromyc
23	160.5	4.2	1585	2 Q9HCC7	Q9HCC7 homo sapien
24	160.5	4.2	1614	2 Q76N89	Q76N89 homo sapien
25	159	4.2	757	2 Q15355	Q15355 homo sapien
26	158.5	4.2	5303	2 Q9V628	Q9V628 drosophila
27	158	4.2	2768	2 Q9VC00	Q9VC00 drosophila
28	157	4.1	2766	2 Q9QZR8	Q9QZR8 rattus norv
29	156.5	4.1	919	2 Q6ZPY9	Q6ZPY9 mus musculus
30	156.5	4.1	1476	2 Q9UP83	Q9UP83 homo sapien
31	156	4.1	746	2 Q9BE82	Q9BE82 macaca fasc

ALIGNMENTS

O60309 homo sapien
Q9C2H4 neurospora
Q90WE3 xenopus lae
Q68DY2 homo sapien
Q81WU2 homo sapien
Q9AY58 oryza sativ
Q9Y6V0 homo sapien
Q6CNF1 kluyveromyc
Q871B2 neurospora
Q6BEP8 homo sapien
Q6BEP8 homo sapien
Q844P8 mus musculu
Q8DEL3 gallus gall
Q9ULJ8 homo sapien
Q9YU36 gallus gall
Q81YP4 homo sapien
Q72BX6 leptospira
Q6SNP9 mus musculu
Q9NSW2 homo sapien
Q89274 mus musculu
Q91X10 mus musculu
Q95153 homo sapien
Q80TB4 mus musculu
Q80TB4 mus musculu
Q8N7U4 homo sapien
Q7R6F5 giardia lam
Q8F2J8 leptospira
Q80TF0 mus musculu
Q35243 mus musculu
Q969M9 homo sapien
Q8R3B7 mus musculu
Q8C049 mus musculu
Q6Z8E1 oryza sativ
Q60732 homo sapien
Q81EB6 plasmodium
Q81R22 drosophila
Q9H986 homo sapien
Q756B4 ashbya goss
Q6CEY9 yarrowia li
Q6GPH8 homo sapien
Q94G15 oryza sativ
Q69ZG6 mus musculu
Q6ZPU4 mus musculu
Q616G8 mus musculu
Q8WU94 homo sapien
Q96SR8 homo sapien
Q67VC7 oryza sativ
Q81WC7 homo sapien
P34926 rattus norv
Q8QNE4 ectocarpus
Q43178 homo sapien
Q6ZMP2 homo sapien
Q72Y5 homo sapien
Q61493 mus musculu
Q8GUP2 caenorhabdi
Q7CS13 mus musculu
Q9BQ11 homo sapien
Q81Y79 homo sapien
Q8EVB9 mycoplasma
Q8TCL4 homo sapien
Q00708 babesia bov
Q91K11 drosophila
Q9VX08 drosophila
Q8DB34 vibrio vuln
Q8C4A5 mus musculu
Q8C4A5 mus musculu
Q7NN14 gloeobacter
Q95JJ5 macaca fasc
Q81J56 plasmodium
Q6PAN9 mus musculu
Q811F2 mus musculu
Q9P2P5 homo sapien

```
RESULT 1
Q6P156 PRELIMINARY; PRT; 713 AA.
AC Q6P156;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Protein phosphatase 1, regulatory subunit 15B.
GN Name=PP1R15B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Director MGC Project;
RA Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065280; AAH65280.1; -.
SQ SEQUENCE 713 AA; 79125 MW; 76C2DA38F9E76A85 CRC64;

Query Match 99.8%; Score 3795; DB 2; Length 713;
Best Local Similarity 99.7%; Pred. No. 1.3e-201;
Matches 711; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEGTGGSRKRLGPRAGFRFPFPPRPSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60
Db 1 MEGTGGSRKRLGPRAGFRFPFPPRPSQAGSKFPTPLGPENSGNPTLLSSAQPTRV 60

Qy 61 SYWTKLLSQLAPLGLLQKVLINSLQFGGMPFTRWLDFAGVYSALRALKRGKRPAAPTA 120
Db 61 SYWTKLLSQLAPLGLLQKVLINSLQFGGMPFTRWLDFAGVYSALRALKRGKRPAAPTA 120

Qy 121 QKSLSSQLDSDSPSVTSPDLWLEGHIIHQVSPDPLKLELKAGSALDPAQAFLLEQOL 180
Db 121 QKSLSSQLDSDSPSVTSPDLWLEGHIIHQVSPDPLKLELKAGSALDPAQAFLLEQOL 180

Qy 181 WGVLLPSSQLRSLYNRELSSPGPLNIQRIIDNFSVYLLNPSYLDLCPRLVSYQN 240
Db 181 WGVLLPSSQLRSLYNRELSSPGPLNIQRIIDNFSVYLLNPSYLDLCPRLVSYQN 240

Qy 241 SDGNSVGVFQTLTPESSCLREDHCHPQLPSAELIPASWQGCPLPSTGLPEIHLRMKR 300
Db 241 SDGNSVGVFQTLTPESSCLREDHCHPQLPSAELIPASWQGCPLPSTGLPEIHLRMKR 300

Qy 301 LEFLQAAKGGDLPTPDQDNGYHSLSEHSLRLMDPKHCRDNPQTFVPAAGDIPGNTQES 360
Db 301 LEFLQAAKGGDLPTPDQDNGYHSLSEHSLRLMDPKHCRDNPQTFVPAAGDIPGNTQES 360

361 TEEKIELLTTEVPLALEEESPSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLPI SARPA 420
361 TEEKIELLTTEVPLALEEESPSEGCPSSEIPMEKEPGEGRISVVVDYSYLEGDLPI SARPA 420

421 CSNKLIIDYILGGASSDLETSSDPGEWDDEAEDDGFDSLSLSDSLDLEQDPGLHLWNS 480
421 CSNKLIIDYILGGASSDLETSSDPGEWDDEAEDDGFDSLSLSDSLDLEQDPGLHLWNS 480

481 FCSVDYPNPONFTATIOAARIIVPEPSPDSEKDLGSKDLENSQSGLPETHSSGEE 540
481 FCSVDYPNPONFTATIOAARIIVPEPSPDSEKDLGSKDLENSQSGLPETHSSGEE 540

541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDSTPSEIVASE 600
541 DDWESSADEAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDSTPSEIVASE 600

601 CHTLLSKVQLLGSQSECPDSVORDVLSGRHTRHVRKKVTFLEEVTEYIISGDERKG 660
601 CHTLLSKVQLLGSQSECPDSVORDVLSGRHTRHVRKKVTFLEEVTEYIISGDERKG 660

661 PWEEFARDGCRFKRIQIETSDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713
661 PWEEFARDGCRFKRIQIETSDAIGYCLTFEHRERMFNRLOQTCFGLNLVKQC 713

RESULT 2
Q96SN1 PRELIMINARY; PRT; 713 AA.
AC Q96SN1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14744.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Yushashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihara K., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK027650; BAB55266.1; -.
```



```
DR GO:0042542; P:response to hydrogen peroxide; IMP.
KW GO:0006979; P:response to oxidative stress; IDA.
FT Hypothetical protein.
FT NON_TER 376
SQ SEQUENCE 376 AA; 41949 MW; 972525D3B472C8F8 CRC64;

Query Match      32.6%; Score 1241; DB 2; Length 376;
Best Local Similarity 65.1%; Pred. No. 8.4e-61;
Matches 255; Conservative 35; Mismatches 86; Indels 16; Gaps 5;

Qy 1 MEPGTGSRKRLGRPRGFRFWPPFPFRSAGSKFPTPLGPENSGNPTLLSSAQPTRV 60
Db 1 METGTHARKRPGRLGSRWFLPFL-RRSHACSEFPFPSSRQPNGN-----SALPERRT 54

Qy 61 SYWTKLSQLLAPLPGLLQKVLINSQFGMFFTRWLDPAFYSAALRAKRGKPAAPTA 120
Db 55 RYWTKLSQLLAPLPSLFQKLLMSQLFGGIPTRWLDPAFYSAALRAKRGKSAAPT 114

Qy 121 QKSLSSQLDSDPSVTPSLDWLEEGHWOYSPDPLKLEKAGSALDPAAQAFLLEQOL 180
Db 115 QKSLSSQLDSDSDLVVSLDWLEEGHWOYSPDPLKLEKAGSALDPAAQAFLLEQOL 174

Qy 181 WGVLLPSSQLSRLYSNRELSSGPGPLNIQRIINFVSVYLLNPSYLDLCPRLVSYQN 240
Db 175 WGVLLPSSQLAGLVSHRELSSGPGPLSVQSLGNFKVSVYLLNPSYLDLCPQLGRQCS 234

Qy 241 SDGNSEVVGFTLTLPSSCLREDHCHPOPLSAELIPASWQCPLSTEGLEPIHHLRMR 300
Db 235 SAGGQGVGFTLTLPSSCYLSEDCGCHPOPLRAEMSAWARRCPPLSTEGLEPIHHLRMR 294

Qy 301 LEFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 295 LEFL-QANKGQELPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 349

Qy 361 TEEKIELTTVPLALEEESSEGCPSSEIPM 392
Db 350 TEKKPELVIEV-----SQSQGSSLFCELPV 376

RESULT 7
Q6VZB6 Q6VZB6 PRELIMINARY; PRT; 158 AA.
AC 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE CNP231 Myd116-like domain protein.
GN Name=CNP231;
OS Canaripox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=44088;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-111;
RX PubMed=14671117; DOI=10.1128/JVI.78.1.353-366.2004;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of canaripox virus.";
RL J. Virol. 78:353-366(2004).
DR EMBL; AY318871; AAR83577.1; -.
SQ SEQUENCE 158 AA; 18892 MW; FBE7CDF3E796421C CRC64;

Query Match      7.0%; Score 267; DB 2; Length 158;
Best Local Similarity 33.7%; Pred. No. 1.9e-07;
Matches 67; Conservative 25; Mismatches 55; Indels 52; Gaps 7;

Qy 505 EEPDSEKDLGKSD--LENSSQSGSLPETPEHSSGEDEDDWESSAD-EAESKLWNSFCN 561
Db 4 QEKADVKKMSLDVNGVENKELS-KLSETDSFQEWDESDSDNDTDTENMLWNMFCR 62

Qy 562 SDDPYNPLNFAPQTSGENKGRDSTKTPSEISVAISECHTLLSCVKQLLGSESECPD 621
Db 63 SDDPYNPLNFTTASVN-----KEMHS-TSCHIDI----- 89

GO:0042542; P:response to hydrogen peroxide; IMP.
KW GO:0006979; P:response to oxidative stress; IDA.
FT Hypothetical protein.
FT NON_TER 376
SQ SEQUENCE 376 AA; 41949 MW; 972525D3B472C8F8 CRC64;

Query Match      32.6%; Score 1241; DB 2; Length 376;
Best Local Similarity 65.1%; Pred. No. 8.4e-61;
Matches 255; Conservative 35; Mismatches 86; Indels 16; Gaps 5;

Qy 1 MEPGTGSRKRLGRPRGFRFWPPFPFRSAGSKFPTPLGPENSGNPTLLSSAQPTRV 60
Db 1 METGTHARKRPGRLGSRWFLPFL-RRSHACSEFPFPSSRQPNGN-----SALPERRT 54

Qy 61 SYWTKLSQLLAPLPGLLQKVLINSQFGMFFTRWLDPAFYSAALRAKRGKPAAPTA 120
Db 55 RYWTKLSQLLAPLPSLFQKLLMSQLFGGIPTRWLDPAFYSAALRAKRGKSAAPT 114

Qy 121 QKSLSSQLDSDPSVTPSLDWLEEGHWOYSPDPLKLEKAGSALDPAAQAFLLEQOL 180
Db 115 QKSLSSQLDSDSDLVVSLDWLEEGHWOYSPDPLKLEKAGSALDPAAQAFLLEQOL 174

Qy 181 WGVLLPSSQLSRLYSNRELSSGPGPLNIQRIINFVSVYLLNPSYLDLCPRLVSYQN 240
Db 175 WGVLLPSSQLAGLVSHRELSSGPGPLSVQSLGNFKVSVYLLNPSYLDLCPQLGRQCS 234

Qy 241 SDGNSEVVGFTLTLPSSCLREDHCHPOPLSAELIPASWQCPLSTEGLEPIHHLRMR 300
Db 235 SAGGQGVGFTLTLPSSCYLSEDCGCHPOPLRAEMSAWARRCPPLSTEGLEPIHHLRMR 294

Qy 301 LEFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 360
Db 295 LEFL-QANKGQELPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQFVPAAGDIPGNTQES 349

Qy 361 TEEKIELTTVPLALEEESSEGCPSSEIPM 392
Db 350 TEKKPELVIEV-----SQSQGSSLFCELPV 376

RESULT 7
Q6IA96 Q6IA96 PRELIMINARY; PRT; 674 AA.
AC 06IA96;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE PP1R15A protein.
GN Name=PP1R15A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR457259; CAG33540.1; -.
SQ SEQUENCE 674 AA; 73523 MW; 88A2EE1AF5DE678 CRC64;

Query Match      7.0%; Score 265.5; DB 2; Length 674;
Best Local Similarity 21.9%; Pred. No. 1.5e-06;
Matches 164; Conservative 81; Mismatches 262; Indels 241; Gaps 32;

Qy 70 LLAPLPGLQKVLINSQFGMFPTR-WLDPAGVYSAL--RALKGR----- 112
Db 21 LLSPVMGLLSRT--WSRL-RGLGPLEPWLVEAVKGAALVEAGLEGEARTPLAIPHTWGR 77

Qy 113 -----EKPAAPTAQKSLSSQLDSDPSVTPSLDWLEEGHWOYSPDPLKLEKAGSA 166
Db 78 RPBEAEADSGGDEDRITLGTSSSLPEAWGLLD-DDGMYGEREATVP---RGQSQ 133

Qy 167 LDPAAQAFLEQOLGWVLELPSLSQSLRYNRELSSGPGPLNIQRIINFVSVYLLNPS 226
Db 134 FADQGRAPL-----SPSLIIRTLQGSQKDPGEKAEEGVAEEGVNKFES----YPPS 182

Qy 227 YLCDFPRLVSYQNSGNSGVGFTLTLPSSCLREDH-CHPOPLSAELIPASWQCPL 285
Db 183 HRECCPAVE---BEDD-----BEAVKKEAHTSTSTALSFGSKPSTWVSCP-- 224

Qy 286 STEGLPEIHLRMRKLEFLQANKGQDLPTPDQDNGVHSLSEEHSLRLMDPKHCRDNPTQ 345
Db 225 ---GEENQATEDKRTERSKGARK-----TSVSPRSSGSDPRS 259

Qy 346 FVPAAGDIPGNTQESTEEKIELTTTEVPLALEEESSEGCPSSEI-----PMEKEPG 397
Db 260 WEYRSG-----EASEEKEKAHKTGKEAAPGPOSSAPAQRLQLKSWMCQPSDEEG 312

Qy 398 EGRISVVDVSYLGGDLPIARPACSNKLIIDYILGGASSDLETSSDP----- 443
Db 313 EVK-----ALGAAEKDGEAECPPCIPPPSAFLKAWVY 344

Qy 444 -EGEDWDERAEDGFDSDSLSDLEQDPEG-----LHLWNSFCSDVPYNQNF 492
Db 345 WPGEDTTEE-EDDEEDSDSGSDEEGEAEASSTPATGVFLKSW-----VYQGED 396

Qy 493 TATIQTAAIRVPEPSDEKDLGKSLGKSLDNSSGSLP-----ETPH-- 535
Db 397 T-----EEDEEDSDTGSADEREAEATSASTPPASAFKAWVYRPGEDTESEED 445

Qy 536 -----SSGEEDWESSADAESKLWNSFCNSDDPYNPLNFKAPQTSG-----ENEKG 584
Db 446 EDVDSKEDDSEALGAEAS-----DPHPSHP-DQRAHFRGMVYRPGKETBEEEA 495
```


Query Match	6.8%;	Score	259.5;	DB 2;	Length	674;			
Best Local Similarity	22.3%;	Pred. No.	3.2e-06;						
Matches	166;	Conservative	82;	Mismatches	261;	Indels	237;	Gaps	34;

QY	70	LLAPLPCLLOKVLWISOLFQGMFPTTR-WLDFAGVYSAL--RAIKGR-----	112
DB	21	LLSPVMGLLSRA--WSRL-RGLGLEPWLVEAVKGAALVEAGLEGEARTPLAIPHTWGR	77
QY	113	-----EKPAAPTAQKSLSLQLDSDSPSVTSPLDWLEEGHIHWQYSPDCLKLELKAUKGSA	166
DB	78	RPGEAEADSGPGEDRETGLTKTSSSLPEAWGLLD-DDDGMYGREATSVP---RGQSGQ	133
QY	167	LDPAQAFLLEQLQWGVLELLPSSLSQSLYNSNRELGSSPGSPGLNTQRIDNFSWVSYLNP	226
DB	134	FADQORAPL-----SPSLIRTLQSGDKPKGEKAEEGVAEEGVNKFSS-----YPPS	182
QY	227	YLCDFPRLEVSYQNSDGNSEVVGQTITPSSCLREDH-CHPQPLSAELIPASWGQCPL	285
DB	183	HRECCPAVE---BEDD-----BEAVKKEAHTSTSAALSPGSKPSTWVSCP--	224
QY	286	STGGLPIHHLRWKRLFLQQAQKQDLPTPDQDNGVYSLSEEHSLLRMPQKHCRRDNPQTQ	345
DB	225	---GEEENQATEKRTKRSKGARK-----TSVSPRSSGSDPRS	259
QY	346	FVPAAGDIPGNTQSTSEKIELLTTEVPLALEEESPSEGC--PSSEIPIPMK-----E	395
DB	260	WEYRSG-----EASEEKEE-----KAHEETGKEAAPGPOSSAPAQORPOLKSWMCQ	305
QY	396	PGGCRISVVDYSYLEGDLPLISARPACSNKLIDYILGCASSDLTSSDP-----	443
DB	306	PSDEEESV-----KP-----LGAAEKDGEAEBCPFCIPPPSAFLKAW	342
QY	444	---EGEDWDREAEDDGDPSDSSLSDSLDLEQDPEG-----LHLWNSFCSDVPDNPQ	490
DB	343	VYVPGEDTEE-EDDEEEDSDSGSDEEGEAASSTPATGVFLKSW-----VYQPG	394
QY	491	NFTATTQTAARIYPEEPSDEKLSGKSDLENSSQSGSLP-----ETPEH	535
DB	395	EDT-----EEDEEDSDTGSABDEREAETSASTPPASAFLLKAWVYRPGEDTEEE	443
QY	536	-----SSGEDDWESSADEAELSLLWNVSFCNSDDPYNPLNFKAPPTQSGENKGCGRD	589

[illegible]

```
QY 232 PRLEVSQNSDGNSEVVGFTLTPESSCLREDHCHQPLSAELIPASWQCPLSTBGLP 291
DB 95 VEAESSPPETWGLSNVDYNA-KGQDDLRKEMERTAGKATLQAGLQAD----- 145
QY 292 EIHILMRKRLFLFOQANKG---QDLPTPDQNGVHSLBEEHSLR-----MDPKHCR 340
DB 146 -----KRLGEVVARBEGVAEPAYPTSOLEGGSPAENEEDGETVKTYYQASAASTAPGYKP 198
QY 341 DNPTQFPAAGDIPGNTQESTEEKIELTTTEVPLALEBEESEPCSPSS--EIPMEKEPGE 398
DB 199 STVPVFLGEA-----EHQATEEK-----GTE-NKADPSNFPSSGSHRAWYYISREKPKQ 247
QY 399 GRISVDVSYLEGDLPTISARPA-----CSNKLID-----YILGGAS 434
DB 248 -----EGEAKVEAHRAAGQGHCPCRNAEABEGGPETTFTVCTGNAFLKAWVYRPGED 296
QY 435 SDLETSSDPE-----CEDWDDEABDDGFDSDSLSDSL 468
DB 297 TEEDNDSDSABEDTATGATPHTSAFLKAWVYRPGEDTEEE-----DSDSDSABEDT 350
QY 469 EQDPEGLHLWNSFCSVPYNPONT-----ATIQTAARI---VPEEP 507
DB 351 AQTGATHT-SAFKAWVYRPGEDTEENS DLSAEDTATGATPHTSAFLKAWVYRPG 409
QY 508 SDSEKOLSGKSDL-----ENSSQSGSLPET-----PEHSSGEEDDWES 545
DB 410 EDTSEE---NSDLSAEDTATGATPHTSAFLKAWVYRPGEDTDDTEEDSEENVAPG 466
QY 546 ---SADBAESLKLWNSFC-----NSDDPYNPLNFKAPFQTSGENEKGCRDKTSPSES 594
DB 467 DSETADSSQSPCLQORCLPGEKTKGRGEEP--PL-FQVAFYLPGEK-----PES 513
QY 595 IVAISECHTLLSKVQLLGS-----QESECSDSVQDVLSCGRHHTVVRKVKVTFLEEVTEY 650
DB 514 PWAAPKLPLRLQRLRUFKAPTRQDDPEIP-----LKARKVHFPAEKVTVH 558
QY 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTPEHRRMNFRLQ 700
DB 559 FLAVWAGPAQAARRGPEQFARDRSFARRIAQAEKGLPYLTIDTSARAWARLR 613

RESULT 12
Q7TQC2 PRELIMINARY; PRT; 578 AA.
AC Q7TQC2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Protein phosphatase regulatory subunit 15A.
GN Name=Gadd34;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=22697643; PubMed=12813455; DOI=10.1038/sj.onc.1206567;
RA Hollander M.C., Poola-Kella S., Fornace A.J. Jr.;
RT "Gadd34 functional domains involved in growth suppression and
apoptosis.";
RL Oncogene 22:3827-3832(2003).
DR EMBL; AY128642; AAM77795.1;
SQ SEQUENCE 578 AA; 63569 MW; 923BC49921C0BC61 CRC64;

Query Match 5.7%; Score 217; DB 2; Length 578;
Best Local Similarity 21.3%; Pred. No. 0.00058;
Matches 152; Conservative 80; Mismatches 225; Indels 258; Gaps 33;

QY 53 SAQPETRVSVYTKLLS-QLLAPLGLQKLVIMQLFGGMPTTWL--DFAGV----- 102
DB 4 SPRPQ-HVLHWKEAHSFLLSPLMGFLSRA--WSRLRGPEVSEAWLAETVAGANQIBADA 60
```

```
QY 103 -----YSALRAKLGREKPA-APTAQKSUSSEQLQSSDPSVTSPLDMLBEGIHQY 151
DB 61 LLTPPPVSENHPLURETEGNTPEWSKAAQRLCLDVEAQSSPKT-----WGL 108
QY 152 SPDPKLKLEKAKGASALDPAQAFLLEQQLWGVELLFSSLOSRLYSNRELSSSGPLNIQ 211
DB 109 SDID-----EHNGK-----PGQDGLRQEVHTHTAGLP-TLQ-----PUHLQ 143
QY 212 RIQNFVSVYLLNPSYLDGFPRLVSVQNSDGNSEVVGFTLTPESSCLREDHCHQPLS 271
DB 144 GADK-----KGVGVVAREEGVSELA----- 163
QY 272 AELIPAS-WQGP-----PLSTEGLPETIHLRMKLEFLQOANKGQDLPTPDQDNGYHSL 327
DB 164 ---YPTSHWEGGPAEDEEDTETVKAH-----QASAAAIAP-----GY----- 198
QY 328 EHSLLRMMDPKHCRDNTQFPVPAAGDIPGNTQESTEEKI-----ELTTTEVPLALE 377
DB 199 -----KPSTSVYVCPGEABHRAATEEKTGDNKAEPSSGSHRWYHTRERPKQEG 246
QY 378 EESFSGCPSSETPMEKEPGEGRISVVVSYLYEGDLPTISARPACSNKLIIDYILGGASDL 437
DB 247 ETXPEQHRAGQSHPCCNAEAE-----EGGPETS---VCS-----GSAFL 282
QY 438 ETSDDPEGEDWDEABDDGFDSDSLSDLEQDPEGLHLWNSFCSVPYNPONTATIQ 497
DB 283 KAWVYRPGEDTEEE-----EDSLDSEED-----TAHTCTTTPH-----TSAFL 321
QY 498 TAARIIVPEPSDSEKDLGKSLDENSQSGSLPET-----PEHSSGEEDDWESSA 547
DB 322 KAWVYRPGEDTEEDDGDWDSAEEDASQCTTHTSAFLKAWVYRPGEDTEEDDSDNSVA 381
QY 548 D-BAESLKLWNSFCNSDDPYNPLNFKAPFQTSGENEKGCRDKTSPSESIV----- 596
DB 382 PVDSETV-----DSCQSTQHCLPV-----EKTGCGEAPFPFQVAFYLPQKQAP 427
QY 597 --AISECHTLLSKVQLL-----GSQSECPDSVQDVLSCGRHHTVVRKVKVTFLEEVTEY 650
DB 428 PWAAPKLPLRLQRLRSLFKAPARNQDPEIP-----LKGRKVFHFSEKTVTVH 472
QY 651 YIS-----GDEDRKGPWEFARDGCRFQKRIQETEDAIGYCLTPEHRRMNFRLQ 700
DB 473 FLAVWAGPAQAARRGPEQFARDRSFARRIAQAEKGLPYLTAPFRARAWTRLR 527

RESULT 13
Q6IN02 PRELIMINARY; PRT; 578 AA.
AC Q6IN02;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Myd116 protein.
GN Name=Myd116;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
```


Job time : 81.7548 secs
